Department of Clinical Sciences & Nutrition

MSc

Exercise and Nutrition Science

Project Title	Physiological responses during performance of the 15-metre Multistage Shuttle Run Test (15mMSFT), with reference to the Police Fitness Standards

Module title	Research Project
Module code	XN7523
Supervisor	Mike Morris

Student J number	J31900
Year of intake	3
Date submitted	13/09/2019

Word count	
Review Paper	4393
Research Article	4385

ii. Declaration

I declare the following project is my own work and have correctly acknowledged the work of others.

Acknowledgments

I would like to thank Mike Morris, Elizabeth Parker and Kevin Sykes for providing me with the opportunity to work alongside them on this unique study, and for their support and guidance throughout the investigation. I would also like to make a special thanks to all those who supported me on throughout this project, Bronwyn Davies, Julia Smith and last but not least Kathryn Manser and Clive Manser. Your support has been invaluable.

iii. Contents page

iv. Abbreviations

Abbreviations	
CVD	Cardiovascular Disease
CRF	Cardiorespiratory Fitness
METs	Metabolic Equivalents
HR	Heart rate
95% CI	95% Confidence Interval
CHD	Coronary Heart Disease
$\dot{V}O_2max$	Maximal Oxygen Consumption
PARE	Physical abilities requirement evaluation
PAT	Physical Appraisal Test
JRFT	Job Related Fitness Test Standard
$L{\cdot}min^{-1}$	Litres of Oxygen per Minute
O_2	Oxygen
$ml{\cdot}kg^{-1}{\cdot}min^{-1}$	Millilitres of Oxygen per Kilogram of body weight per Minute
MSFT	Multi-stage Fitness Test
$\dot{V}O_2$	Oxygen Consumption
$\dot{V}CO_2$	Carbon Dioxide production
RER	Respiratory Exchange Ratio
RPE	Ratings of Perceived Exertion
SEE	Standard Error of Estimate
ICC	Intra-Class Correlation Coefficient
EPOC	Exercise Post Oxygen Consumption
LoA	Limits of Agreement
$m{\cdot}s^{-1}$	meters per second
$km{\cdot}h^{-1}$	kilometres per hour

v. Tables and Figures

Literature review	
Table and Figures	**Description**
Table 1	Recommended standards (level: shuttle) and estimated aerobic capacity (ml·kg^{-1}·min^{-1}) of specialist police groups. Adapted from Brewer et al. (2010).
Table 2	$\dot{V}O_2$ peak values (ml·kg^{-1}·min^{-1}) of boys, girls and total sample from the criterion method and 4 predictive equations (n=115). Adapted from Batista et al. (2013).
Figure 1	Percentage increase in muscle activation at Pre-Contact (PC), Weight Acceptance (WA), and Peak Push Off (PP0) with a Crossover cut technique (XOV), 60° (S60) or 30° (S30) angle change during the sidestepping techniques. The asterixis (*) indicates a significant difference compared to straight line technique. Adapted from Besier et al. (2003).
Table 4	Distance of shuttles and turn frequency corresponding to a running velocity at 4.3 and 5.4 km·h^{-1}. Adapted from Hatamoto et al. (2013).
Table 5	The Mean oxygen cost of change in direction at various velocities. Adapted from Hatamoto et al. (2014).

Research Project	
Table and Figures	**Description**
Table 1	Table 1. Police fitness standards for 14 police roles including the Recommended Standard (level:shuttle) and Estimated aerobic capacity (ml·kg^{-1}·min^{-1}) adapted from Brewer et al. (2010).
Table 2	The 15-metre MSFT protocol adapted from Bleep test (2019).
Table 3	Physiological Responses at level 5:4; 7:6; 9:4 and 10:5 for Trial One and Comparison to Estimated Aerobic Capacity Required.
Table 4	Reliability analyses of measured $\dot{V}O_2$ between trial 1 and 2 displaying 95% Limits of Agreement (LoA), Intraclass correlation coefficient (ICC) and Typical Error.
Table 5	Measured $\dot{V}O_2$ at level 5:4 and 7:6 for males and females.
Table 6	The estimated aerobic and anaerobic energy contribution at $\dot{V}O_2max$ upon level 5:4, 7:6, 9:4 and 10:5 of the 15-metre MSFT, based on the findings of Bertuzzi et al. (2013).

Abstract

The objective of this review is to provide a broad outline of the research surrounding the validity and reliability of the 15-metre multi-stage fitness test (MSFT) for measuring the aerobic fitness of police officers. Maintenance of optimal cardiorespiratory fitness (CRF) in the emergency services is vital for health maintenance, injury prevention, and physical preparation for on-duty tasks.

Police officers in England and Wales are required to attend annual fitness testing with minimal standards in place for entry into police safety training (PST). The current minimal standard is level 5:4, an estimated $\dot{V}O_2max$ of $35ml\cdot kg^{-1}\cdot min^{-1}$, with the requirements increasing for specialist roles. This is assessed using the 15-metre MSFT previously developed and validated against laboratory obtained measures. Previous validation studies have compared the physiological responses between the 15-metre MSFT and training protocols for varying police roles.

For example, Brewer, Buckle & Castle (2013) validated the level 5:4 standard by assessing the heart rate responses between the PST and level 5:4 of the 15-metre MSFT. Despite greater peak heart rate responses reported in the 15-metre MSFT (Peak heart rate: $175\pm13\ b\cdot min^{-1}$ vs. $152\pm12\ b\cdot min^{-1}$), the standard was maintained with concerns the aerobic fitness of police officers would be suboptimal for the role and below that of the general population. Using a similar methodology, the minimal entry requirements for 13 additional roles were developed and validated.

However to date, the validity and reliability of the 15-metre MSFT has not been assessed using direct measures of gas analysis, previously relying on indirect measures to assess demands.

Keywords: Cardiovascular Disease, Cardiorespiratory Fitness, Maximal Oxygen Uptake, Multistage Fitness Test, Validity, Reliability, Shuttle Running, Constant Speed Running.

1. Literature review

1.1. Introduction- Importance of occupational health and fitness

The risk of occupational death is profoundly higher within the emergency services (Zimmerman, 2012), in 2014 it was estimated occupational death among United States police officers was 13.5 per 100,000 deaths compared to the national average of 3.4 (Bureau of Labour Statistics, 2016).

Whilst the majority of deaths are related to on-duty tasks (Tiesman, Hendricks, Bell & Amandus, 2010), cardiovascular disease (CVD) related events are estimated to contribute to 7% of sudden cardiac deaths among police officers (Zimmerman, 2012) and 45% among firefighters (Kales, Soteriades, Christoudias & Christiani, 2003; Smith, Barr & Kales, 2013). Previous research has highlighted that on-duty cardiovascular events are more prevalent among those with CVD risk factors or underlying diseases (Baur, Christophi, Tsismenakis, Cook, Kales, 2011; Holder, Stallings, Peeples, Burress & Kales, 2006; Kales et al. 2003).

Cardiorespiratory fitness (CRF), defined as “the ability of the circulatory, respiratory, and muscular systems to supply oxygen during sustained physical activity” (Lee, Artero, Sui & Blair, 2010), is a factor associated with CVD event risk and premature mortality (Gander et al. 2015; Kodama et al. 2009). Within the general population, epidemiological studies have established increasing CRF or maintaining high fitness levels are protective against CVD related conditions and all-cause mortality (Gander et al. 2015; Laukkanen et al. 2001; Prestgaard et al. 2019; Kodama et al. 2009). In a 22 year follow up (1979-2002), Gander et al. (2015) reported men in the highest CRF category (13.6 Metabolic equivalents (METs)), exhibited a 33% reduction (Hazard Ratio: 0.67: 95% CI: 0.51-0.88) in Coronary Heart Disease (CHD) development than those in the low CRF category (8.6 METs). In a 32-35 year follow up amongst 1403 males (Age: 40-59), Prestgaard et al. (2019) estimated men that maintained high fitness levels or those who became fit reduced the risk of all-cause mortality by 35-43% (Hazard Ratio: 0.57: 95% CI: 0.17-0.67; Hazard Ratio: 0.65: 95% CI: 0.46–0.90, respectively) compared to those who became unfit. In contrast, men with high fitness levels at baseline which declined at follow up almost doubled the risk of all-cause mortality (Hazard Ratio: 1.74: 95% CI: 1.35-2.23) (Prestgaard et al. 2019). Kodama et al. (2009) estimated a 1 MET (3.5 $ml \cdot kg^{-1} \cdot min^{-1}$) increase in aerobic capacity resulted in significantly lowered risk of both CHD/CVD events by 15% (0.85, 95% CI, 0.82-0.88, $p<0.001$) and all-cause mortality by 13% (Hazard Ratio: 0.87: 95% CI: 0.84-0.90, $p<0.001$).

Whilst high CRF levels are protective against CVD related events and all-cause mortality, it may also reduce the risk of injury and sickness absence amongst emergency service personnel (Poplin, Roe, Peate, Harris & Burgess, 2014; Wynn & Hawdon, 2012; Kyröläinen et al. 2008). Poplin et al. (2014) examined the risk of injury among 773 firefighters, comparing aerobic fitness levels and risk of sustaining a job-related injury over a 4-year period (2005-2009). It was estimated firefighters in the lowest CRF category (<43$ml \cdot kg^{-1} \cdot min^{-1}$) had a 2.2 times higher risk of sustaining any type of injury (Hazard Ratio: 2.22: 95% CI: 1.72-2.88), compared to the highest category (>48 ml·kg^{-1}·min^{-1}) (Poplin et al. 2014). In the median percentile (43–48 ml·kg^{-1}·min^{-1}), this risk was increased by almost 40% compared to the highest fitness category (Hazard Ratio: 1.38: 1.38-1.78)) (Poplin et al. 2014). It was further estimated that injury risk was reduced by 14% when aerobic capacity was improved by one MET (3.5 ml·kg^{-1}·min^{-1}) (Poplin et al. 2014). Kyröläinen et al. (2008) analysed data collected from 7179 male military personnel on sickness absence and fitness profiles. Participants with 7+ days sickness in a year had significantly lower ($p<0.001$) $\dot{V}O_2max$ scores compared to both those with 1-7 days or no absence (Combinate $\dot{V}O_2max$: 44.7±9.1 vs. 46.2±8.2 and 47.1±7.9 ml·kg^{-1}·min^{-1}, respectively) (Kyröläinen et al. 2008).

In addition to CRF contributing to protection against adverse health conditions/all-cause mortality, it also plays an important role in job compatibility.

For police officers, optimal CRF levels along with muscular power, strength and endurance, and flexibility are essential for on-the-job responsibilities (Lockie, Dawes, Kornhauser & Holmes, 2019; Crawley, Sherman, Crawley & Cosa-Lima, 2016; Dawes, Orr; Siekaniec, Vanderwoude & Pope, 2016). Whilst this occupation generally involves a high proportion of low-intensity tasks/activities such as patrolling by car or on foot (Bonneau & Brown, 1995), a small proportion requires police officers to perform a variety of short burst high-intensity movements including sprinting, jumping and apprehending of suspects (Anderson, Plecas & Segger, 2001; Bonneau & Brown, 1995).

The infrequent nature of the job makes it imperative that police officers are physically prepared to make quick transitions between sedentary or low-intensity tasks, to potentially hostile situations which may require maximal exertion (Crawley et al. 2015;

Bonneau & Brown, 1995). Screening protocols are necessary to recruit candidates that can effectively respond to these situations (Anderson et al. 2001; Strating, Bakker, Dijkstra, Lemmink & Groothoff, 2010). Failure to screen out incompetent candidates may endanger public safety, cause long-term injury or disability risk, reduce productivity and increase employee turnover, impacting both human and economic cost (Anderson et al. 2001; Strating et al. 2010; Kyröläinen et al. 2008).

Physical examination of police and emergency services is common across multiple nations, including the Canadian police requiring completion of the PARE test (Physical abilities requirement evaluation) (http://www.rcmp-grc.gc.ca/en/prepare-for-pare) and the New Zealand police using the physical appraisal test (PAT) (https://www.newcops.co.nz/start-training/pat-training-advice).

In the UK, current serving police officers and new applicants are required to complete the Job-related fitness test standard (JRFT) on an annual basis for entry into police safety training (PST) (College of Policing, 2014). This requires attainment of level 5:4 of the 15-metre MSFT (College of Policing, 2014). In addition, minimal entry requirements are in place for specialist police roles requiring attainment of up to level 10:5 on the 15-metre MSFT (Brewer, Morgan, Breivik, Dyer & Harmer, 2010).

1.2 Origin of the 15-metre multi-stage fitness test

Typically, aerobic fitness of individuals is examined in laboratory conditions accurately assessing a variety of key physiological variables, classically determined by a graded test to exhaustion using a motorised treadmill or cycle ergometer (Pate & Kriska, 1984; Castagna, Impellizzeri, Charmi, Carlomagno & Rampinini, 2006). During this test, pulmonary gas measurements are taken including $\dot{V}O_2$ described in absolute terms as the litres of oxygen per minute ($L \cdot min^{-1}$), or relatively taking into account body mass as millilitres of oxygen (O_2) per kilogram of body weight per minute ($ml \cdot kg^{-1} \cdot min^{-1}$) (Powers & Howley, 2018). The purpose of this measure is to determine Maximal oxygen uptake ($\dot{V}O_2max$), the “maximal capacity to transport and utilize oxygen during exercise” (Powers & Howley, 2018, p.76).

Whilst accuracy is major advantage to this form of testing, it requires expensive equipment and trained staff which can be time consuming for both the examiner and participants involved (Castagna et al. 2006). As alternatives, field-based tests have been developed to estimate $\dot{V}O_2max$, such as the 20-metre multistage-fitness test (20-metre MSFT) which has been validated against values reported in a laboratory setting (Léger, Mercier, Gadoury & Lambert, 1988; Léger & Goudary, 1989; Ramsbottom, Brewer & Williams, 1988). The minimalist equipment required, minimal training required from supervisors, and potential for multiple participants being tested at one time is a major advantage of this test format (Léger & Lambert, 1982; Léger & Goudary, 1989). The test requires participants to run between cones marked 20 metres apart starting at 8.5 $km \cdot h^{-1}$, increasing by 0.5 $km \cdot h^{-1}$ every minute (Léger et al.

1988). A pre-recorded CD provides audio cues (bleeps) indicating the start and end of a shuttle/stage (Léger et al. 1988). The test is ended either by voluntary cessation or failure to maintain the pace set, determined by falling three metres behind the line at the "bleep" (Léger et al. 1988).

From 2013, police officers and new recruits within the UK are required to meet the JRFT standards in accordance with recommendations by the Police Advisory Board for England and Wales (College of Policing, 2014). This requires minimal attainment of level 5:4 on the 15-metre MSFT, a predicted $\dot{V}0_2max$ of 35 $ml{\cdot}kg^{-1}{\cdot}min^{-1}$, completed annually (College of Policing, 2014). The 15-metre MSFT was originally designed by the Loughborough University of Technology (Hazeldine, Lakomy, Simpson, 1995). This was adapted from the 20-metre MSFT to a 15-metre distance to accommodate limited space at assessment centres (Brewer et al. 2010). The original standard for this test was attainment of level 8.1 (42.5 $ml{\cdot}kg^{-1}{\cdot}min^{-1}$) to match the average aerobic capacity ($\dot{V}O_2max$) of the general population between 12-54 years of age, which has since been altered (Hazeldine et al. 1995).

The Roehampton University (Roehampton, 2003, unpublished study) is thought to have first developed the level 5:4 standard. A total of 40 participants (36.3±9.7 years) completed the 15-metre MSFT and a maximal graded treadmill test three days apart (Roehampton, 2003, unpublished study). The data obtained from each test enabled backward extrapolation (via linear regression) between directly measured $\dot{V}O_2max$ on a maximal graded treadmill test, and number of shuttles completed during the 15-metre MSFT (Roehampton, 2003, Unpublished study). It was found the current minimal standard (level 6:5) overestimated the true value found from the maximal treadmill test (Level 6:5 estimate: 34.7 $ml \cdot kg^{-1} \cdot min^{-1}$; Actual measurement: 37.71 $ml \cdot kg^{-1} \cdot min^{-1}$) (Roehampton, 2003, Unpublished study). Thus, to match 34.7$ml \cdot kg^{-1} \cdot min^{-1}$, the standard was lowered to level 5:4. However the study provided no information on confidence intervals or error margins, therefore it was unclear if bias or random error effected these results.

Using level 5:4 as a minimal standard, Brewer, Buckle & Castle (2013) conducted a validation study comparing both heart rate (HR) and blood lactate responses of male (n=75) and female (n=44) police officers (31.7±5.3 years) completing the 15-metre MSFT and "Officer safety training" (OST). The OST was a four-hour course including role play activities and simulations reflecting situations police officers may face on patrol (Brewer et al. 2013). After completing the 15-metre MSFT, a 10-15-minute recovery period was allowed before starting the OST (Brewer et al. 2013).

Using an independent samples t-test, both the peak and mean HR was found to be significantly higher in the 15-metre MSFT compared to the OST (Peak: 175±13 $b \cdot min^{-1}$ vs. 152±12 $b \cdot min^{-1}$, $p<0.05$; Mean: 158±11 $b \cdot min^{-1}$ vs. 126±10 $b \cdot min^{-1}$, $p<0.05$, respectively) (Brewer et al. 2013). Simarly post-exercise blood lactate was found to be higher in the 15-metre MSFT compared to the OST (4.5±2.1 $mmol \cdot l^{-1}$ vs. 2.8±0.9 $mmol \cdot l^{-1}$, $p<0.05$) (Brewer et al. 2013). The author commented whilst the physiological responses in the MSFT exceeded that found in the OST, some individuals expressed HR close to that experienced in the MSFT (Brewer et al. 2013). Therefore, concluding that lowering the standards below 35 $ml.kg^{-1}min^{-1}$ could result in officers with sub-optimal aerobic fitness to cope with the peak demands of the MSFT and below that of a normal population, therefore supporting level 5:4 as a sufficient minimal standard (Brewer et al. 2013).

A noted limitation of the study by Brewer et al. (2013) was direct measures of oxygen uptake ($\dot{V}O_2$) were not taken during the MSFT, instead relying on indirect measures such as HR as an alternative marker of $\dot{V}O_2$, based on the relationship between $\dot{V}O_2$, HR and Work rate (Arts & Kuipers, 1994; Lamberts, Lemmink, Durandt & Lambert, 2004). Measured HR during prediction of $\dot{V}O_2$ is sensitive to external factors including temperature and humidity, hydration status, caffeine consumption, medication use, and emotional state which may influence the HR-$\dot{V}O_2$ relationship (Astrand & Rodahl, 1986; Powers & Howley, 2018). Additionally, HR can lag behind changes in movement patterns, remaining elevated after ending the movement (Trost, 2007), thereby potentially limiting accuracy of measurement.

Secondly, with only a recovery period of 10-15 minutes between activities $\dot{V}O_2$ may remain elevated after cessation of exercise known as the exercise post oxygen consumption (EPOC), resultant of metabolic processes related to recovery from exercise including replenishment of oxygen and phosphocreatine stores and increased body temperature (Power & Howley, 2018). Some have reported EPOC can last several minutes (Brehm & Gutin, 1986; Maresh et al. 1992) to several hours (Maehlum, Grandmontagne, Newsholme, Sejersted, 1986; Tomlin & Wenger, 2001) dependant on the intensity of exercise undertaken. Additionally, HR can remain elevated above pre-exercise levels for up to 60 minutes after moderate to vigorous exercise (Takahashi, Okada, Hayano & Tamura, 2002; Takahashi, Okada, Saitoh, Hayano & Miyamoto, 2000; Brown, Chitwood, Anderson & Boatwright, 1993; Van Hooren & Peake, 2018). Therefore, both $\dot{V}O_2$ and HR may have been elevated above resting levels upon starting the OST, thereby potentially causing bias in the results.

Prior to the research reported by Brewer et al. (2013), Brewer et al. (2010) utilised a similar methodology comparing HR responses of the 15-metre MSFT and simulated training protocols. The results enabled the development of 13 additional specialist group standards ranging from level 5:4 up to 10:5 (51 $ml \cdot kg^{-1} \cdot min^{-1}$) (Table 1). However, comparable to Brewer et al. (2013), used heart responses to validate the standards with no information of data spread, therefore limiting the overall analysis.

Table 1. Recommended standards (level: shuttle) and estimated aerobic capacity ($ml \cdot kg^{-1} \cdot min^{-1}$) of specialist police groups. Adapted from Brewer et al. (2010).

Unit	Recommended standard (level: shuttle)	Estimated aerobic capacity ($ml \cdot kg^{-1} \cdot min^{-1}$)
PST	5: 4	35
Marine police unit	5: 4	35
CBRN	5: 4	35
Method of entry	5: 4	35
Dog handler	5: 7	36
Mounted branch	5: 7	36
Police cyclist	5: 8	36
Police support unit	6: 3	37
Air support	6: 4	37
Police divers	6: 8	39
Marine police (tactical skills)	7: 2	40
Authorized firearms officer	7: 6	41
Armed response vehicle	9: 4	46
Dynamic intervention AFO	10: 5	51

1.3 Validity and reliability of the multistage fitness test

To gain an understanding of the degree of measurement error of protocols or devices, two important factors should be considered; co-current validity and retest reliability (Hopkins, 2000). Validity describes the ability of a measurement tool or protocol to measure what it is intended to do (Currell & Jeukendrup, 2008; Atkinson & Nevill, 1998), whilst co-current validity is described as “the agreement between the observed value and the true or criterion value of a measure” (Hopkins, 2000, p.1).

Retest reliability is described as “reproducibility of the observed value when the measurement is repeated” (Hopkins, 2000, p.1), with Atkinson & Nevill, (1998) considering measurement error as being associated with two components of variability: systematic bias and random error. Systematic bias/change describes the difference in the mean between trials either in a positive or negative direction (Atkinson & Nevill, 1998; Hopkins, 2000). An example of this is the learning effect; a subject may improve performance upon retest due to a desire to improve or advancement in skill indicating a positive direction of systematic bias (Atkinson & Nevill, 1998; Hopkins, 2000). Conversely, random error is the error that occurs during the sampling process, including mechanical and biological differences and protocol inconsistencies (e.g. changes in technique between exercise trials) (Atkinson & Nevill, 1998).

1.3.1 Validity

Whilst investigations into the validity and reliability of the 15-metre MSFT are limited, a similar field-based method of assessing aerobic fitness, the 20-metre MSFT, has been widely researched investigating the validity of predictive equations (Stickland, Petersen & Bouffard, 2003; Ramsbottom et al. 1988; Barnett, Chan & Bruce, 1993; Léger et al. 1988; Mahar, Welk, Rowe, Crotts & Mclver, 2006; Matsuzaka et al. 2004). Stickland et al. (2003) examined the validity of two predictive equations (Léger et al. 1988; Léger & Goudary, 1989) to predict the $\dot{V}O_2max$ obtained during a graded treadmill test. Men (n=60) and women (n=62) (Age: 18-38 years) achieved a mean $\dot{V}O_2max$ of 54.9±8.4 $ml{\cdot}kg^{-1}{\cdot}min^{-1}$ and 47.4±6.2 $ml{\cdot}kg^{-1}{\cdot}min^{-1}$ and a mean shuttle run (SR) level of 9.5±2.7 and 7.8±1.8, respectively (Stickland et al. 2003). A Pearson's-correlation coefficient revealed strong and significant correlations between SR and $\dot{V}O_2max$ for men and women (r=0.88, r=0.81, p<0.01, respectively) (Stickland et al. 2003). An independent samples t-test revealed both equations significantly underestimated $\dot{V}O_2max$ by ~3.3-4.2 $ml{\cdot}kg^{-1}{\cdot}min^{-1}$ for men (Léger et al. (1988): 51.6±8.1, p<0.05; Léger & Gadoury (1989): 50.7±8.9, p<0.05) and by ~1-2.3 $ml{\cdot}kg^{-1}{\cdot}min^{-1}$ with women (Léger et al. (1988): 46.4±5.3, p<0.05; Léger & Gadoury (1989): 45.1±5.8, p<0.05) (Stickland et al. 2003).

Kilding, Aziz & Teh (2006) compared the predicted $\dot{V}O_2max$ from two equations (Léger et al. 1988; Ramsbottom et al. 1988) with directly measured oxygen uptake obtained during the 20-metre MSFT among 26 male athletes (age: 21.9±3.5 years). Compared to directly measured $\dot{V}O_2max$ (59.1±6.6 $ml{\cdot}kg^{-1}{\cdot}min^{-1}$), the predictive equations resulted in a significant underestimation by 9.3% (Léger et al. 1988: 53.6±3.9 $ml{\cdot}kg^{-1}{\cdot}min^{-1}$) and 13.2% (Ramsbottom et al. 1988: 51.3±4 $ml{\cdot}kg^{-1}{\cdot}min^{-1}$) (Kilding et al. 2006). The

Pearson's product-moment correlation indicated moderate and significant relationships between measured and predicted values ($r=0.61-0.68$, $p<0.05$) (Kilding et al. 2006).

Batista et al. (2013) assessed the validity of four predictive equations (Barnett et al. 1993; Léger et al. 1988; Mahar et al. 2006; Matsuzaka et al. 2004) among 115 adolescents (Boys: 12.3±0.9; Girls 12.1±0.7 years). The repeated measures analysis of variance (ANOVA) revealed all equations significantly underestimated directly measured $\dot{V}O_2max$ on a graded treadmill test (Criterion method) by 2.0-5.6 $ml \cdot kg^{-1} \cdot min^{-1}$ ($p<0.05$) (Table 2), except for the Mahar et al. (2006) equation for boys accounting for body mass, gender and number of laps ($p>0.05$) (Batista et al. 2013).

Table 2. $\dot{V}O_2$ peak values ($ml \cdot kg^{-1} \cdot min^{-1}$) of boys, girls and total sample from the criterion method and 4 predictive equations (n=115). Adapted from Batista et al. (2013).

Variables	$\dot{V}O_2$ peak values ($ml \cdot kg^{-1} \cdot min^{-1}$)		
	Boys (n=61)	Girls (n=54)	Total (n=115)
Criterion method	49.9±9.5	42.2±7.4	46.3±9.4
Barnett et al. (1993)	47.8±4.5*	40.3±3.8*	44.3±5.6*
Léger et al. (1988)	41.6±4.2*	39.7±3.1*	40.7±3.9*
Mahar et al. (2006)	50.2±5.1	43.8±4.9*	47.2±5.6
Matsuzaka et al. (2004)	43.7±6.3*	40.7±5.0*	42.3±5.9*

*p<0.05 vs. reference method.

Studies investigating the MSFT have used statistical methods including the Pearson's correlation coefficients and/or hypothesis testing including the paired samples t-test or repeated measures ANOVA to evaluate validity or reliability (Batista et al. 2013; Stickland et al. 2003; Kilding et al. 2006). The Pearson's correlation coefficient predominantly indicates the relationship between measures not agreement (Cooper, Baker, Tong, Roberts & Hanford, 2005; Atkinson & Nevill, 1998; Berchtold, 2016), and is sensitive to the spread (heterogeneity) of data, therefore individual differences may affect the overall relationship between measures (Hopkins, 2000).

The paired samples t-test and repeated measures ANOVA detects bias between repeat measures, with a paired samples t-test comparing the means, and the latter comparing multiple test-retest results utilising post hoc tests to detect systematic bias (Atkinson & Nevill, 1988). However, these methods are unable to detect random error, therefore relying on these tests alone are not recommended (Atkinson & Nevill, 1998).

1.3.2 Potential reasons for discrepancy between the criterion and estimated values

It has been suggested the differences between directly measured $\dot{V}O_2max$ from the criterion method and predictive equations could be attributed to the differences in exercise mode (Ruiz et al. 2009). The criterion method (the graded treadmill test) requires straight-line running, whilst shuttle running requires frequent changes of direction (Ruiz et al. 2009). Previous research comparing these running techniques unequivocally report greater energy costs with shuttle running, demonstrating a 30-50% increase in energy costs (Stevens et al. 2015; Zamparo, Zadro, Lazzer, Beato & Sepulcri, 2014; Buglione & di Prampero, 2013). This will be covered in further detail in section 1.4. When the predictive equations were developed, a regression equation was used correlating laboratory obtained $\dot{V}O_2max$ during the maximal graded treadmill test and an indirect measure of performance during the MSFT (i.e. shuttles/levels completed) (Ramsbottom et al. 1988; Léger et al. 1988; Léger & Goudary, 1989)

Due to the differences in energy cost between the running techniques, it could be speculated the model may not account for additional oxygen/energy costs required with shuttle running causing the differences between predicted and measured $\dot{V}O_2max$ (Ruiz et al. 2009; Stojanovic et al. 2016; Silva et al. 2012). To address this, studies have developed and validated predictive equations based on direct measures of $\dot{V}O_2$ taken during the 20-metre MSFT (Ruiz et al. 2009; Silva et al. 2012).

Silva et al. (2012) examined an equation based on directly measured $\dot{V}O_2max$ during the 20-metre MSFT accounting for Gender, Body Mass Index and shuttle run stage. Among 54 girls and boys 10-18 years, no mean differences (MD) ($p>0.05$) were present between directly measured $\dot{V}O_2max$ during the 20-metre MSFT (48.1±9.5ml.kg^{-1}·min^{-1}) and the equation (MD: 0.0ml.kg^{-1}·min^{-1}; 95% LoA: 11.2-11.2). When compared to the Léger et al. (1988) equation developed from laboratory obtained $\dot{V}O_2max$, there was a larger mean difference between directly measured $\dot{V}O_2max$ (Estimated $\dot{V}O_2max$: 45.2±5.9ml.kg^{-1}·min^{-1}; MD: -2.9ml.kg^{-1}·min^{-1}; 95% LoA: -16.8; 11.0) (Silva et al. 2012). Despite the improved accuracy of the new equation, the author highlighted random error (95% LoA) affected all predictive equations and the possibility of result bias should be considered with future research (Silva et al. 2012). In contrast, Ruiz et al. (2009) reported an equation based on directly measured $\dot{V}O_2max$ during the 20-metre MSFT (Ruiz et al. 2008), this significantly underestimated the $\dot{V}O_2max$ of male and female youth (Age: 14-15 years) obtained during the 20-metre MSFT (-3.7±1.7 ml·kg^{-1}·min^{-1}) with a standard error of estimate (SEE) between measures of ±5.3ml·kg^{-1}·min^{-1}.

Whilst there is some suggestion of improved prediction of $\dot{V}O_2$max using equations based on direct measures as found with Silva et al. (2012), research is limited using this methodology, yielding equivocal results. Therefore, further research is required to affirm these findings. Furthermore, with these equations being based on only youth and adolescents to date, further research focusing on adult participants may be required.

1.3.3 Reliability

The reliability of the 20-metre MSFT has been assessed using a variety of statistical methods, with early research typically correlating the shuttles completed between trials (Liu, Plowman & Looney, 1992; Léger et al. 1988; Vincent, Barker, Clarke & Harrison, 1999; Sproule Kunalan, McNeil, Wright, 1993; Mahoney, 1992).

Liu et al. (1992) examined the test-retest reliability of the 20-metre MSFT among 12 males (n=12) and females (n=8), testing participants on two separate occasions 1-week apart. The Pearson's intra-class correlation-coefficient (ICC) indicated a high correlation between the mean number of laps (shuttles) (r=0.93) completed between trial 1 and 2 (Liu et al. 1992). The mean laps completed between trials was not significantly different (T1: 47.80±20.3; T2: 50.55±22.4, p>0.05) (Liu et al. 1992). Several other authors reported similarly high test-retest correlations with children and adults (r=0.73-0.95) (Léger et al. 1988; Vincent et al. 1999; Sproule et al. 1993;

Mahoney, 1992), however, the mean or error margins for the performance measures correlated (i.e. shuttles completed) was not reported.

Similar to validity, the Pearson's correlations coefficient will only indicate the strength of a relationship between scores not the agreement and provides no information of absolute variation between scores (Atkinson & Nevil, 1998; Lamb & Rogers, 2007). When assessing the reliability of data, it is recommended to include measures of absolute reliability due to the method being unaffected by the range of measurements (Atkinson & Nevill, 1998). The recommended technique by Atkinson & Nevill (1998), is the Bland & Altman (1986) 95% Limits of Agreement (LoA) due to the ability to detect trial-to-trial random error. The 95% LoA presents the individual differences between measures against the mean of the measures, which indicates a range where 95% of the differences lie between measures (Bland & Altman, 1986; Bai, Shu & Niu, 2019). Observing a narrow range would indicate a close agreement between measures, whilst a broad range would be indicative of poor agreement (Bai et al. 2019). Atkinson & Nevill (1998) noted that the 95% LoA may be held if heteroscedasticity is near zero and the differences between measures are normally distributed.

Several studies investigating the 20-metre MSFT have utilised this statistical method to coincide with relative reliability measures (Aandstand, Holme, Berntsen & Anderssen, 2011; Copper et al. 2005; Lamb & Rogers, 2007).

Aandstad et al. (2011) examined the test-retest differences among male soldiers (n= 38, Age: 34.8±4.0 years), performing the 20-metre MSFT on two separate occasions. Using a newly developed equation to predict $\dot{V}O_2max$, the paired samples t-test indicated a significant difference ($p<0.05$) between trial 1 and 2 estimated $\dot{V}O_2max$ (49.8±5.0 vs. 50.6±4.7 $ml\cdot kg^{-1}\cdot min^{-1}$, respectively), with the 95% LoA demonstrating mean bias and random error between trials of -0.8±3.1 $ml\cdot kg^{-1}\cdot min^{-1}$ (Aandstad et al. 2011).

Cooper, Baker, Tong, Roberts & Hanford (2005) examined the reliability of the 20-metre MSFT among 21 active men (21±3.6 years) completing the test on two separate occasions. The predicted $\dot{V}O_2max$ obtained from trial 1 and 2 indicated non-significant bias between trials (Trial 1: 52.9±8.8 $ml\cdot kg^{-1}\cdot min^{-1}$; Trial 2: 53.3±8.9 $ml\cdot kg^{-1}\cdot min^{-1}$, $p=0.190$) and a 95% LoA of -0.4±2.7 $ml\cdot kg^{-1}\cdot min^{-1}$ (Cooper et al. 2005). Both authors reported the reliability to be acceptable due to the conservative mean bias and random error (Cooper et al. 2005; Aandstad et al. 2011).

Lamb & Rogers (2007) examined reliability of the 20-metre MSFT with 22 males (20.9±1.5 years) and 13 females (19.6±1.0 years) on three separate occasions one-week apart, using the Léger et al. (1988) equation to predict $\dot{V}O_2max$. A repeated measure ANOVA revealed $\dot{V}O_2max$ for trial 2 (53.3±8.4 $ml\cdot kg^{-1}\cdot min^{-1}$) and trial 3 (52.9±7.6 $ml\cdot kg^{-1}\cdot min^{-1}$) was significantly ($p<0.05$) higher than trial 1 (52.1±7.8 $ml\cdot kg^{-1}\cdot min^{-1}$). The 95% LoA analysis demonstrated a diminishment in mean bias with -1.1±4.7 $ml\cdot kg^{-1}\cdot min^{-1}$ between trial 1 and 2, reducing to 0.0±5.0 $ml\cdot kg^{-1}\cdot min^{-1}$ between trial 2 and 3 (Lamb & Rogers, 2007).

Overall the reliability of the 20-metre MSFT is acceptable, demonstrating narrow mean bias and variance (random error) between measures. However to date, reliability has only been investigated using predictive equations. Future research is warranted to investigate reliability using direct measures of gas analysis.

1.4 Shuttle vs. constant speed running

As mentioned earlier, the differences in running techniques between the criterion (straight-line running) and 20-metre MSFT (Shuttle-Running) may contribute to the disparities in predicted $\dot{V}O_2max$. The energy cost differences between these techniques have previously been highlighted (Stevens et al. 2015; Buglione & di Prampero, 2013; Fessi, Farhat, Dellal, Malone, Moalla, 2018; Tang et al. 2018). Stevens et al. (2015) investigated the measured oxygen cost of continuous 10-metre shuttle running and constant-speed running (160m track without change of direction) using breath-by-breath gas analysis using a portable gas analyser (K4b2; Cosmed Slr, Rome, Italy). Among fourteen male soccer players (23±2 years; $\dot{V}O_2max$: 54±6 $ml \cdot kg^{-1} \cdot min^{-1}$) shuttle running elicited a significantly higher energy cost by 30-50% when compared to constant-speed running at matched average speeds (Stevens et al. 2015). For example, at an approximate speed of 10.0 $km \cdot h^{-1}$ (constant-speed: 2.78 $m \cdot s^{-1}$; Shuttle: 2.56 $m \cdot s^{-1}$), constant speed running demonstrated a measured energy cost of 4.60±0.26 $J \cdot kg^{-1} \cdot m^{-1}$, whilst shuttle running displayed a higher cost of 6.71±0.45 $J \cdot kg^{-1} \cdot m^{-1}$ (37.3% difference) (Stevens et al. 2015).

Zamparo et al. (2014) also reported a greater energy/oxygen cost of shuttle running. During this study, participants performed an incremental treadmill test (Constant-speed running) and shuttle-running of varying running speeds and shuttle lengths. Shuttle-running with a 10-metre distance (180° turn angle) demonstrated a substantially higher energy/oxygen cost (Net energy cost: 22.1±0.9 $J \cdot m^{-1} \cdot kg^{-1}$; $\dot{V}O_2$: 33.9±0.6 $ml \cdot kg^{-1} \cdot min^{-1}$) than constant speed running (Net energy cost: 3.97±0.34 $J \cdot kg^{-1} \cdot min^{-1}$) (Zamparo et al. 2014). However, energy costs of constant-speed running were based on a speed of 2.77 $m \cdot s^{-1}$ (10 $km \cdot h^{-1}$) over a 4-minute period opposed to an average speed of 4.51±0.20 $m \cdot s^{-1}$ with shuttle running (Zamparo et al. 2014).

A noted limitation with studies comparing shuttle and constant-speed running is velocity (speed) is not often matched (Zamparo et al. 2014; Zadro, Sepulcri, Lazzer, Fregolent & Zamparo, 2011), therefore this may not provide a true representation of the metabolic cost differences. The additional energy/oxygen costs of shuttle running are thought to be due to differences in technique, with shuttle running requiring frequent change of direction (Hatamoto et al. 2013; Hatamoto et al. 2014; Buchheit, Haydar, Hader, Ufland & Ahmaidi, 2011) and increased muscle activation in acceleration and deceleration phases (Besier, Lloyd & Ackland, 2003). Besier et al. (2003) investigated muscle activation patterns during straight-line running and cutting manoeuvres (cross-over technique and 30/60-degree change of direction) that were both pre-planned and unanticipated. In pre-planned conditions, according to Electromyographic data, all three cutting manoeuvres caused significant increases in average muscle activation during precontact (PC), weight acceptance (WA) and peak push off (PPO) phases (Figure 1), all significantly greater than straight-line running ($p<0.01$) (Beiser et al. 2003). The greatest muscle activation was observed with a greater direction change with an average activation of ~65%, 35% and 35% with 60° direction change in the respective movement phases, and ~35%, 20% and 25% with 30° direction change (Besier et al. 2003). Overall the results suggested muscle activation was significantly higher than straight-line running, and the greater angle used in the cutting technique further increased muscle activation (Besier et al. 2003). This may therefore contribute to the additional energy costs involved in shuttle running.

Figure 1. Percentage increase in muscle activation at Pre-Contact (PC), Weight Acceptance (WA), and Peak Push Off (PP0) with a Crossover cut technique (XOV), 60° (S60) or 30° (S30) angle change during the sidestepping techniques. The asterixis (*) indicates a significant difference compared to straight line technique. Adapted from Besier et al. (2003).

Additional factors including shuttle length, approach speed to the turn, and frequency of turns could also influence the energy/oxygen cost of shuttle running (Hatamoto et al. 2013; Hatamoto et al. 2014; Buglione & di Prampero, 2013; Minetti, Ardigò, Capodaglio & Saibene, 2001; Minetti, Gaudino, Seminati & Cazzola, 2013).

Buglione & di Prampero (2013) examined directly measured energy cost (via K4b^2 portable gas analyser) of physically active males (n=65; 23.7±7.3 years) performing 10 and 20-metre shuttle runs. Findings indicated significantly higher mean energy cost of 10-metre than 20-metre shuttles at 2.86m·s^{-1} (6.88±0.19 vs. 5.32±0.13 J·m^{-1}·kg^{-1}, p<0.05) (Buglione & di Prampero, 2013). Simarly at 4.00 $m \cdot s^{-1}$, the mean energy cost was higher with 10-metre shuttles almost doubling the energy cost of 20-metre shuttles (14.29±0.75 $J \cdot m^{-1} \cdot kg^{-1}$ vs. 7.52±0.26 $J \cdot m^{-1} \cdot kg^{-1}$, p<0.05), therefore suggesting augmented energy costs at shorter distances and greater running velocities (Buglione & di Prampero, 2013).

Hatamoto and colleagues (Hatamoto et al. 2013; Hatamoto et al. 2014) attempted to quantify the energy cost of a change of direction and the impact of turn frequency. During the first experiment, participants completed shuttles at fixed speeds of 4.3 or 5.4 $km \cdot h^{-1}$ between 3-9 metres, causing 8-24 turns or 10-30 turns per minute for 4.3 and 5.4$km \cdot h^{-1}$ respectively (Table 3) (Hatamoto et al. 2013). Among ten active males (22.8±2 years), a significant positive correlation was found between the mean $\dot{V}O_2$ and turn frequency for both 4.3$km \cdot h^{-1}$ (r=0.973, p<0.05) and 5.4$km \cdot h^{-1}$ (r=0.996, p<0.05) indicating a higher oxygen cost was required for higher turn frequencies (Hatamoto et al. 2013). Additionally, the oxygen cost of turning was significantly greater at a velocity of 5.4$km \cdot h^{-1}$ compared to 4.3 $km \cdot h^{-1}$ (0.34±0.13 $ml \cdot kg^{-1} \cdot min^{-1}$ [95% CI: 0.193–0.492] vs. 0.55±0.09 $ml \cdot kg^{-1} \cdot min^{-1}$ [95% CI: 0.467–0.641], respectively, p<0.001), with the physiological demand of a single turn estimated to be 7.2±12.0 $J \cdot kg^{-1}$ and 12.0±2.1 $J \cdot kg^{-1}$ for 4.3 and 5.4 $km \cdot h^{-1}$ respectively (Hatamoto et al. 2013).

Using a similar methodology, Hatamoto et al. (2014) also found the energy cost of a change of direction proliferated as velocity increased at 3, 4, 5, 6,7 and 8 $km \cdot h^{-1}$ (Table 4), with differences between velocities being significant (p<0.05) except between 7 and 8 $km \cdot h^{-1}$ (p=0.110). Additionally, change of direction frequency had a substantial effect on energy cost when compared to a matched speed on the treadmill. For example, a velocity of 3km·h^{-1} with 30 turns per minute had over a 50% higher gross $\dot{V}O_2$ cost than straight-line treadmill running (22.7±1.8 vs. 13.4±1.0 ml·kg^{-1}·min^{-1}), with a velocity double that on the treadmill (6km·h^{-1}) being required to produce a similar metabolic cost (22.0±1.2 ml·kg^{-1}·min^{-1}) (Hatamoto et al. 2014).

Table 3. Distance of shuttles and turn frequency corresponding to a running velocity at 4.3 and 5.4 km·h^{-1}. Adapted from Hatamoto et al. (2013).

Velocity	Distance of shuttle and turn frequency (turns per minute)				
	3 metres	3.6 metres	4.5 metres	6 metres	9 metres
4.3 (km·h^{-1})	24	20	16	12	8
5.4 (km·h^{-1})	30	25	20	15	10

Table 4. The Mean oxygen cost of change in direction at various velocities. Adapted from Hatamoto et al. (2014).

Velocity (km·h^{-1})	**Oxygen cost (ml·kg^{-1}·min^{-1})**
3	0.27±0.03
4	0.35±0.05
5	0.48±0.10
6	0.68±0.08
7	0.87±0.13
8	0.99±0.14

1.5. Conclusion

The studies looked upon in this review detail the use of the 15-metre MSFT in assessing the aerobic fitness of police officers. Despite the research done in these studies, there are limitations in their methodology to validate police fitness standards.

The 20-metre MSFT, similar to the 15-metre MSFT, uses predictive equations/tables based on backward extrapolation techniques which results in underestimations of $\dot{V}O_2max$ in comparison to criterion (maximal graded treadmill performance). As noted in section 1.3.2 and 1.4, there is speculation that the discrepancy between methods is due to the greater energy costs required with shuttle running in comparison to straight-line running. This then poses the question of whether the additional energy costs required for shuttle running have been accounted for in the equations. Whilst previous research has attempted to address this issue by developing the equations based on direct measures during activity, the results are equivocal requiring investigation into adult populations.

Reliability investigations into the 20-metre MSFT has interpreted the test as acceptable for measuring $\dot{V}O_2max$. However, the reliability in relation to direct measures of gas analysis has not been investigated warranting further research into this area.

The research leading to the development of police entry standards has employed a methodology of comparing the physiological responses (i.e. HR and blood lactate) between the 15-metre MSFT and PST. The limitations with these indirect approaches to establish the oxygen cost requirements highlights the potential for these responses to poorly represent the oxygen cost required.

To date, the validity of the estimated aerobic capacity required for police entry levels developed by Brewer et al. (2010) has not been investigated using direct gas analysis. Secondly, the reliability of the test has not been established. Therefore, further research is warranted to investigate if the estimated aerobic capacity from level 5:4 to 10:5 is consistent with direct measures of oxygen uptake during the 15-metre MSFT and to establish the reliability of the test.

1.6. **References**

Aandstad, A., Holme, I., Berntsen, S., & Anderssen, S. A. (2011). Validity and reliability of the 20 meter shuttle run test in military personnel. *Military Medicine, 176* (5), 513-518. doi:10.7205/MILMED-D-10-00373.

Anderson, G. S., Plecas, D., & Segger, T. (2001). Police officer physical ability testing - re-validating a selection criterion. *Policing: An International Journal of Police Strategies & Management, 24*(1), 8-31. doi:10.1108/13639510110382232.

Arts, F. J., & Kuipers, H. (1994). The relation between power output, oxygen uptake and heart rate in male athletes. *International Journal of Sports Medicine*, *15*(5), 228-231.

Astrand, P.O. and Rodahl, K. (1986) *Textbook of work physiology*: *Physiological bases of exercise*. (3rd Ed.). New York, NY. McGraw-Hill. doi:10.2310/6640.2004.00030.

Atkinson, G., Nevill, A, M. (1998) Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Medicine*. *26*(4), 217-38.

Bai, Z., Shu, T., & Niu, W. (2019). Test-retest reliability and measurement errors of grip strength test in patients with traumatic injuries in the upper extremity: A cross-sectional study. *BMC Musculoskeletal Disorders*, *20*(1), 256-256. doi:10.1186/s12891-019-2623-z.

Barnett, A., Chan, L. Y. S., & Bruce, L. C. (1993). A preliminary study of the 20-m multistage shuttle run as a predictor of peak VO2 in Hong Kong Chinese students. *Paediatric Exercise Science*, *5*(1), 42-50. doi:10.1123/pes.5.1.42.

Batista, M. B., Cyrino, E. S., Arruda, M., Dourado, A. C., Coelho-E-Silva, M. J., Ohara, D., . . . Ronque, E. R. V. (2013). Validity of equations for estimating VO2peak from the 20-m shuttle run test in adolescents aged 11–13 years. *Journal of Strength and Conditioning Research*, *27*(10), 2774-2781. doi:10.1519/JSC.0b013e3182815724.

Baur, D. M., Christophi, C. A., Tsismenakis, A. J., Cook, E. F., & Kales, S. N. (2011). Cardiorespiratory fitness predicts cardiovascular risk profiles in career firefighters. *Journal of Occupational and Environmental Medicine*, *53*(10), 1155-1160.

Berchtold, A. (2016). Test–retest: Agreement or reliability? *Methodological Innovations, 9*, 1-7. doi:10.1177/2059799116672875.

Besier, T. F., Lloyd, D. G., & Ackland, T. R. (2003). Muscle activation strategies at the knee during running and cutting maneuvers. *Medicine and Science in Sports and Exercise, 35*(1), 119-127. doi:10.1097/00005768-200301000-00019.

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet (London, England), 1*(8476), 307-310.

Bonneau, J., Brown, J. (1995). Physical ability, fitness and police work. *Journal of Clinical Forensic Medicine. 2*(3), 157-164.

Brehm, B. A., & Gutin, B. (1986). Recovery energy expenditure for steady state exercise in runners and nonexercisers. *Medicine and Science in Sports and Exercise*, 18(2), 205-210. doi:10.1249/00005768-198604000-00010.

Brewer, J., Buckle, P., Castle, P. (2013). The use of workplace physiological measurements to establish the minimum fitness standards required for entry into the united kingdom police service. *Journal of Athletic Enhancement, 2*(4) doi:10.4172/2324-9080.1000121.

Brewer, J., Morgan, N., Breivik, S., Dyer, A & Harmer, D. (2010). *Job Related Fitness Tests for Police Officer Specialist Posts*. UK: National Police Improvement Agency. Retrieved from the College of policing website: https://www.college.police.uk/What-we-do/Standards/Fitness/Documents/FWG_Report_final.pdf.

Brown, S. P., Li, H., Chitwood, L. F., Anderson, E. R., & Boatwright, D. (1993). Blood pressure, hemodynamic, and thermal responses after cycling exercise. *Journal of Applied Physiology, 75*(1), 240-245.

Buchheit, M., Haydar, B., Hader, K., Ufland, P., & Ahmaidi, S. (2011). Assessing running economy during field running with changes of direction: Application to 20 m shuttle runs. *International Journal of Sports Physiology and Performance*, *6*(3), 380-395. doi:10.1123/ijspp.6.3.380.

Buglione, A., & di Prampero, P. E. (2013). The energy cost of shuttle running. *European Journal of Applied Physiology*, *113*(6), 1535-1543. doi:10.1007/s00421-012-2580-9.

Bureau of Labour Statistics. (2016). *Injuries, Illnesses and Fatalities*. Retrieved September 05, 2019 from https://www.bls.gov/iif/oshwc/cfoi/police-officers-2014.htm.

Castagna, C., Impellizzeri, F. M., Chamari, K., Carlomagno, D., & Rampinini, E. (2006). Aerobic fitness and yo-yo continuous and intermittent tests performances in soccer players: A correlation study. *Journal of Strength and Conditioning Research*, 20(2), 320-325. doi:10.1519/R-18065.1.

College of Policing (2014) *Implementation. Job-Related Fitness Tests for the Police Service of England and Wales*: Winsor Recommendation 33. Retrieved from https://www.college.police.uk/What-we-do/Standards/Fitness/Documents/PoliceFitnessGuidance.pdf.

Cooper, S., Baker, J. S., Tong, R. J., Roberts, E., & Hanford, M. (2005). The repeatability and criterion related validity of the 20 m multistage fitness test as a predictor of maximal oxygen uptake in active young men. *British Journal of Sports Medicine*, *39*(4), 1-7. doi:10.1136/bjsm.2004.013078.

Crawley, A. A., Sherman, R. A., Crawley, W. R., & Cosio-Lima, L. M. (2016). Physical fitness of police academy cadets: Baseline characteristics and changes during a 16-week academy. *Journal of Strength and Conditioning Research*, *30*(5), 1416-1424. doi:10.1519/JSC.0000000000001229.

Currell, K., & Jeukendrup, A. E. (2008). Validity, reliability and sensitivity of measures of sporting performance. *Sports medicine, 38* (4), 297-316. doi:10.2165/00007256-200838040-00003.

Dawes, J. J., Orr, R. M., Siekaniec, C. L., Vanderwoude, A. A., & Pope, R. (2016). Associations between anthropometric characteristics and physical performance in male law enforcement officers: A retrospective cohort study. *Annals of Occupational and Environmental Medicine, 28*(1), 26-26. doi:10.1186/s40557-016-0112-5.

Fessi, M. S., Farhat, F., Dellal, A., Malone, J. J., & Moalla, W. (2018). Straight-line and change-of-direction intermittent running in professional soccer players. *International Journal of Sports Physiology and Performance*, *13*(5), 562-567. doi:10.1123/ijspp.2016-0318.

Gander, J. C., Sui, X., Hébert, J. R., Hazlett, L. J., Cai, B., Lavie, C. J., & Blair, S. N. (2015). Association of cardiorespiratory fitness with coronary heart disease in asymptomatic men. *Mayo Clinic Proceedings*, *90*(10), 1372-1379. doi:10.1016/j.mayocp.2015.07.017.

Hatamoto, Y., Yamada, Y., Fujii, T., Higaki, Y., Kiyonaga, A., & Tanaka, H. (2013). A novel method for calculating the energy cost of turning during running. *Journal of Sports Medicine, 9*(4), 117-122. doi:10.2147/OAJSM.S39206.

Hatamoto, Y., Yamada, Y., Sagayama, H., Higaki, Y., Kiyonaga, A., & Tanaka, H. (2014). The relationship between running velocity and the energy cost of turning during running. *PloS One*, *9*(1), e81850-e81850. doi:10.1371/journal.pone.0081850.

Hazeldine, R., Lakomy, H, K, A., & Simpson, D. (1995) *Recommendations for the Standards of Fitness of Metropolitan Police Officers*. Unpublished research, Department of Physical Education, Sports Science and Recreation Management, Loughborough University of Technology, UK.

Holder, J. D., Stallings, L. A., Peeples, L., Burress, J. W., & Kales, S. N. (2006). Firefighter heart presumption retirements in massachusetts 1997-2004. *Journal of Occupational and Environmental Medicine, 48*(10), 1047-1053.

Hopkins, W.G. (2000). Measures of reliability in sports medicine and science. *Sports Medicine, 30*(1), 1-15. doi:10.2165/00007256-200030010-00001.

Kales, S. N., Soteriades, E. S., Christoudias, S. G., & Christiani, D. C. (2003). Firefighters and on-duty deaths from coronary heart disease: A case control study. *Environmental Health: A Global Access Science Source, 2*(1), 1-13. doi:10.1186/1476-069X-2-14.

Kilding, A. E., Aziz, A. R., & Teh, K. C. (2006). Measuring and predicting maximal aerobic power in international-level intermittent sport athletes. T*he Journal of Sports Medicine and Physical Fitness*, *46*(3), 366-372.

Kodama, S., Saito, K., Tanaka, S., Maki, M., Yachi, Y., Asumi, M., . . . Sone, H. (2009). Cardiorespiratory fitness as a quantitative predictor of all-cause mortality and cardiovascular events in healthy men and women: A meta-analysis. *Journal of the American Medical Association*, *301*(19), 2024-2035.

Kyröläinen, H., Häkkinen, K., Kautiainen, H., Santtila, M., Pihlainen, K., & Häkkinen, A. (2008). Physical fitness, BMI and sickness absence in male military personnel. *Occupational Medicine, 58*(4), 251-256.

Lamb, K. L., & Rogers, L. (2007). A re-appraisal of the reliability of the 20 m multi-stage shuttle run test. *European Journal of Applied Physiology*, *100*(3), 287-292. doi:10.1007/s00421-007-0432-9.

Lamberts, R. P., Lemmink, K. A. P. M., Durandt, J. J., & Lambert, M. I. (2004). Variation in heart rate during submaximal exercise: Implications for monitoring training. Journal of *Strength and Conditioning Research*, 18(3), 641-645. doi:10.1519/00124278-200408000-00044.

Laukkanen, J. A., Lakka, T. A., Rauramaa, R., Kuhanen, R., Venäläinen, J. M., Salonen, R., & Salonen, J. T. (2001). Cardiovascular fitness as a predictor of mortality in men. *Archives of Internal Medicine, 161*(6), 825-831.

Lee, D., Artero, E. G., Sui, X., & Blair, S. N. (2010). Mortality trends in the general population: The importance of cardiorespiratory fitness*. Journal of Psychopharmacology, 24(4 Suppl),* 27-35. doi: 10.2147/OAJSM.S39206.

Léger, L. A., & Lambert, J. (1982). A maximal multistage 20-m shuttle run test to predict VO2 max. *European Journal of Applied Physiology and Occupational Physiology, 49*(1), 1-12.

Léger, L. A., Mercier, D., Gadoury, C., & Lambert, J. (1988). The multistage 20 metre shuttle run test for aerobic fitness. *Journal of Sports Sciences, 6*(2), 93-101.

Léger, L., & Gadoury, C. (1989). Validity of the 20 m shuttle run test with 1 min stages to predict VO2max in adults. *Canadian Journal of Sport Sciences, 14*(1), 21-26.

Liu, N. Y., Plowman, S. A., & Looney, M. A. (1992). The reliability and validity of the 20-meter shuttle test in American students 12 to 15 years old. *Research Quarterly for Exercise and Sport, 63*(4), 360-365. doi:10.1080/02701367.1992.10608757.

Lockie, R. G., Dawes, J. J., Kornhauser, C. L., & Holmes, R. J. (2019). Cross-sectional and retrospective cohort analysis of the effects of age on flexibility, strength endurance, lower-body power, and aerobic fitness in law enforcement officers. *Journal of Strength and Conditioning Research, 33*(2), 451-458. doi:10.1519/JSC.0000000000001937.

Maehlum, S., Grandmontagne, M., Newsholme, E. A., & Sejersted, O. M. (1986). Magnitude and duration of excess postexercise oxygen consumption in healthy young participants. *Metabolism: Clinical and Experimental, 35*(5), 425-429.

Mahar, M. T., Welk, G. J., Rowe, D. A., Crotts, D. J., & McIver, K. L. (2006). Development and validation of a regression model to estimate VO2peak from PACER 20-m shuttle run performance. *Journal of Physical Activity and Health*, 3(s2), S34-S46. doi:10.1123/jpah.3.s2.s34.

Mahoney, C. (1992). 20-MST and PWC170 validity in non-Caucasian children in the UK. *British Journal of Sports Medicine, 26*(1), 45-47. doi:10.1136/bjsm.26.1.45.

Maresh, C. M., Abraham, A., De Souza, M. J., Deschenes, M. R., Kraemer, W. J., Armstrong, L. E., . . . Hoffman, J. R. (1992). Oxygen consumption following exercise of moderate intensity and duration. *European Journal of Applied Physiology and Occupational Physiology, 65*(5), 421-426.

Matsuzaka, A., Takahashi, Y., Yamazoe, M., Kumakura, N., Ikeda, A., Wilk, B., & Bar-Or, O. (2004). Validity of the multistage 20-M shuttle-run test for Japanese children, adolescents, and adults. *Pediatric Exercise Science*, *16*(2), 113-125. doi:10.1123/pes.16.2.113.

Minetti, A. E., Ardigò, L. P., Capodaglio, E. M., & Saibene, F. (2001). Energetics and mechanics of human walking at oscillating Speeds1. *American Zoologist, 41*(2), 205-210. doi:10.1668/0003-1569(2001)041[0205:EAMOHW]2.0.CO;2.

Minetti, A. E., Gaudino, P., Seminati, E., & Cazzola, D. (2013). The cost of transport of human running is not affected, as in walking, by wide acceleration/deceleration cycles. *Journal of Applied Physiology, 114*(4), 498-503. doi:10.1152/japplphysiol.00959.2012.

Pate, R. R., & Kriska, A. (1984). Physiological basis of the sex difference in cardiorespiratory endurance. *Sports Medicine, 1*(2), 87-98.

Poplin, G. S., Roe, D. J., Peate, W., Harris, R. B., & Burgess, J. L. (2014). The association of aerobic fitness with injuries in the fire service. *American Journal of Epidemiology*, *179*(2), 149-155.

Powers, S. K., & Howley, E. T. (2018). *Exercise physiology: Theory and application to fitness and performance (Tenth ed.).* New York: McGraw-Hill.

Prestgaard, E., Mariampillai, J., Engeseth, K., Erikssen, J., Bodegård, J., Liestøl, K., . . . Berge, E. (2019). Change in cardiorespiratory fitness and risk of stroke and death: Long-term follow-up of healthy middle-aged men. *Stroke*, 50(1), 155-161. doi:10.1161/STROKEAHA.118.021798.

Ramsbottom, R., Brewer, J., & Williams, C. (1988). A progressive shuttle run test to estimate maximal oxygen uptake. *British Journal of Sports Medicine, 22*(4), 141-144. doi:10.1136/bjsm.22.4.141.

Roehampton University of Surrey. (2003 unpublished study). *Report on the Validation of the 15m Work Related Fitness Test*. UK: Roehampton University of Surrey.

Ruiz, J. R., Ramirez-Lechuga, J., Ortega, F. B., Castro-Piñero, J., Benitez, J. M., Arauzo-Azofra, A., . . . Zabala, M. (2008). Artificial neural network-based equation for estimating VO2max from the 20m shuttle run test in adolescents. *Artificial Intelligence in Medicine*, *44*(3), 233-245. doi:10.1016/j.artmed.2008.06.004.

Ruiz, J. R., Silva, G., Oliveira, N., Ribeiro, J. C., Oliveira, J. F., & Mota, J. (2009). Criterion-related validity of the 20-m shuttle run test in youths aged 13-19 years. *Journal of Sports Sciences, 27*(9), 899-906. doi:10.1080/02640410902902835.

Silva, G., Oliveira, N. L., Aires, L., Mota, J., Oliveira, J., & Ribeiro, J. C. (2012). Calculation and validation of models for estimating VO2max from the 20-m shuttle run test in children and adolescents. *Archives of Exercise in Health and Disease*, 3, 145-152. doi:10.5628/aehd.v3i1-2.20.

Smith, D. L., Barr, D. A., & Kales, S. N. (2013). Extreme sacrifice: Sudden cardiac death in the US fire service. *Extreme Physiology & Medicine*, *2*(1), 2-6. doi:10.1186/2046-7648-2-6.

Sproule, J., Kunalan, C., McNeill, M., & Wright, H. (1993). Validity of 20-MST for predicting VO2max of adult Singaporean athletes. *British Journal of Sports Medicine*, *27*(3), 202-204. doi:10.1136/bjsm.27.3.202.

Stevens, T. G. A., de Ruiter, C. J., van Maurik, D., van Lierop, C. J. W., Savelsbergh, G. J. P., & Beek, P. J. (2015). Measured and estimated energy cost of constant and shuttle running in soccer players. *Medicine and Science in Sports and Exercise, 47*(6), 1219-1224. doi:10.1249/MSS.0000000000000515.

Stickland, M. K., Petersen, S. R., & Bouffard, M. (2003). Prediction of maximal aerobic power from the 20-m multi-stage shuttle run test. *Canadian Journal of Applied Physiology*, *28*(2), 272-282. doi:10.1139/h03-021.

Stojanovic, M. D., Calleja-Gonzalez, J., Mikic, M., Drid, P., Vuckovic, I., & Ostojic, S. M. (2016). Accuracy and criterion-related validity of the 20-M shuttle run test in well-trained young basketball players. *Montenegrin Journal of Sports Science and Medicine, 5*(2), 5-10.

Strating, M., Bakker, R. H., Dijkstra, G. J., Lemmink, K A P M, & Groothoff, J. W. (2010). A job-related fitness test for the dutch police. *Occupational Medicine (Oxford, England), 60*(4), 255-260.

Takahashi, T., Okada, A., Hayano, J., & Tamura, T. (2002). Influence of cool-down exercise on autonomic control of heart rate during recovery from dynamic exercise. *Frontiers of Medical and Biological Engineering: The International Journal of the Japan Society of Medical Electronics and Biological Engineering*, *11*(4), 249-259.

Takahashi, T., Okada, A., Saitoh, T., Hayano, J., & Miyamoto, Y. (2000). Difference in human cardiovascular response between upright and supine recovery from upright cycle exercise. *European Journal of Applied Physiology, 81*(3), 233-239. doi:10.1007/s004210050036.

Tang, R., Murtagh, C., Warrington, G., Cable, T., Morgan, O., O'Boyle, A., . . . Drust, B. (2018). Directional change mediates the physiological response to high-intensity shuttle running in professional soccer players. *Sports (Basel, Switzerland)*, *6*(2), 1-11. doi:10.3390/sports6020039.

Tiesman, H. M., Hendricks, S. A., Bell, J. L., & Amandus, H. A. (2010). Eleven years of occupational mortality in law enforcement: The census of fatal occupational injuries, 1992-2002. *American Journal of Industrial Medicine*, *53*(9), 940-949. doi:10.1002/ajim.20863.

Tomlin, D.L., & Wenger, H.A. (2001). The relationship between aerobic fitness and recovery from high intensity intermittent exercise. *Sports Medicine, 31*(1), 1-11. doi:10.2165/00007256-200131010-00001.

Trost, S. G. (2007). State of the art reviews: Measurement of physical activity in children and adolescents. *SAGE Publications*, 1 (4), 299-314. doi:10.1177/1559827607301686.

Van Hooren, B., & Peake, J. M. (2018). Do we need a cool-down after exercise? A narrative review of the psychophysiological effects and the effects on performance, injuries and the long-term adaptive response. *Sports Medicine*, 48(7), 1575-1595. doi:10.1007/s40279-018-0916-2.

Vincent, S. D., Barker, R., Clarke, M., & Harrison, J. (1999). A comparison of peak heart rates elicited by the 1-mile Run/Walk and the progressive aerobic cardiovascular endurance run. *Research Quarterly for Exercise and Sport*, 70(1), 75-78. doi:10.1080/02701367.1999.10607733.

Wynn, P., & Hawdon, P. (2012). Cardiorespiratory fitness selection standard and occupational outcomes in trainee firefighters. *Occupational Medicine (Oxford, England), 62*(2), 123-8.

Zadro, I., Sepulcri, L., Lazzer, S., Fregolent, R., & Zamparo, P. (2011). A protocol of intermittent exercise (shuttle runs) to train young basketball players. *Journal of Strength and Conditioning Research*, *25*(6), 1767-1773. doi:10.1519/JSC.0b013e3181da85d1.

Zamparo, P., Zadro, I., Lazzer, S., Beato, M., & Sepulcri, L. (2014). Energetics of shuttle runs: The effects of distance and change of direction*. International Journal of Sports Physiology and Performance, 9*(6), 1033-1039. doi:10.1123/ijspp.2013-0258.

Zimmerman, F. H. (2012). Cardiovascular disease and risk factors in law enforcement personnel: A comprehensive review. *Cardiology in Review*, *20*(4), 159-166.

2. Research Project

Physiological responses during performance of the 15-metre Multistage Shuttle Run Test (15mMSFT), with reference to the Police Fitness Standards

2.1. Journal proposal

The research article in question is intended for the Journal of Occupational Health, with focus on the current developments surrounding occupational health. Health promotion will be the primary field investigated of the six subjects covered in the journal. This article will aim to expand on both the knowledge and understanding of a commonly used occupational fitness test within police forces in England and Wales, for the promotion of the health and safety of workers.

2.2. Abstract

Background- Police officers need to meet minimal aerobic fitness standards for entry into specific police roles. The 15-metre multistage-fitness test (MSFT) is used to assess aerobic fitness, however the estimated aerobic capacity required for the entry levels has not been validated using directly measured oxygen uptake. **Aims**- To investigate the validity and reliability at four entry levels (Level: Shuttle: 5:4; 7:6; 9:4; 10:5) of the 15m Multistage Fitness Test (MSFT). **Methods**- Data was collected from 64 UK police officers (41±7 years; 177.1±9.2cm; 82.3±13.8 kg) using a repeated measures design. Officers completed the 15-metre MSFT equipped with a portable gas analyser (Metamax 3B, Cortex, Germany) measuring $\dot{V}O_2$, heart rate and respiratory exchange ratio (RER) throughout testing. Test validity was assessed making comparisons between predicted and absolute $\dot{V}O_2$ at level 5:4, 7:6, 9:4 and 10:5 using a One-Sample T-Test. Between trial test-retest reliability was determined using a Paired Sample t-test, Ninety-five Limits of Agreement (95% LoA), Intra-Class Correlation Coefficient (ICC) and Typical Error. **Results-** The mean difference between predicted and measured $\dot{V}O_2$ at 35 (5:4), 41 (7:6), 46 (9:4) and 51 (10:5) $ml{\cdot}kg^{-1}{\cdot}min^{-1}$ were as followed 5.43, 5.82, 6.07 and 4.32 $ml{\cdot}kg^{-1}{\cdot}min^{-1}$, indicating a significant underpredictions at all levels ($p<0.05$). Reliability analyses indicated no significant bias between trials ($p>0.05$). The 95% LoA analysis revealed the most conservative bias and random error was -0.29±7.7 $ml{\cdot}kg^{-1}{\cdot}min^{-1}$. No gender differences were observed ($p>0.05$). **Conclusion-** The current findings support the use of the minimal estimated aerobic capacities ($\dot{V}O_2max$) required to achieve level 5:4, 7:6, 9:4 and 10:5 of the 15-

metre MSFT. The 15-metre MSFT is generally reliable for assessing $\dot{V}O_2$, with further investigations required into high-fit populations.

Keywords: Physical Fitness Standards, Aerobic Fitness, Multistage Fitness Test, Validity, Reliability, Oxygen Cost, Anaerobic Metabolism, Respiratory Exchange Ratio.

3. Introduction

Physical fitness standards are integral to ensure police officers can meet the day-to-day physical demands of the occupation (Anderson, Plecas & Segger, 2001; Bonneau & Brown, 1995). Whilst a typically sedentary occupation (Bonneau & Brown, 1995), police officers may encounter physically demanding tasks including pursuits on foot or restraint of suspects (Anderson et al. 2001; Bonneau & Brown, 1995). Incompetence may risk public safety or cause self-harm from injury (Anderson et al. 2001; Strating, Bakker, Dijkstra, Lemmink & Groothoff, 2010; Kyröläinen et al. 2008).

Traditionally, aerobic fitness is assessed on the graded treadmill or cycling ergometer where direct measurements of maximal oxygen uptake ($\dot{V}O_2$max) are taken (Pate & Kriska, 1984; Castagna, Impellizzeri, Charmi, Carlomagno & Rampini, 2006). Whilst superior with accuracy and repeatability, access to a laboratory, specialised equipment and trained operators limits viability (Castangna et al. 2006). Alternative field-based methods, such as the 20-metre multistage fitness test (MSFT), have been developed to predict $\dot{V}O_2$max without the use of specialist apparatus (Léger & Lambert, 1982). Following the 2013 Winsor report (Winsor, 2013), police officers in England and Wales are required to complete an adapted 15-metre MSFT to monitor aerobic fitness, to a minimal level of four shuttles at level five (level 5:4) (estimated $\dot{V}O_2$max: 35 ml·kg^{-}

^{1}·min^{-1}) (College of Policing, 2014), a test initially developed by the University of Loughborough in 1995 (Hazeldine, Lakomy, Simpson, 1995) for the metropolitan police accommodating for limited space at assessment centres (Brewer, Morgan, Breivik, Dyer & Harmer, 2010).

Several studies have attempted to validate this test for use within the police (Brewer et al. 2004; Brewer et al. 2010; Brewer, Buckle & Castle, 2013), the first known conducted by the University of Surrey (Roehampton, 2003, Unpublished). By plotting the number of shuttles completed in the 15-metre MSFT against direct measures of $\dot{V}O_2max$, it was found previous minimal standards overpredicted absolute values by approximately 3 ml·kg^{-1}·min^{-1}, thereby reducing the standard from five shuttles at level six to four shuttles at level five (Roehampton, 2003, Unpublished).

Following this, Brewer and colleagues (Brewer et al. 2004; Brewer et al. 2010; Brewer et al. 2013) validated the level 5:4 standard during the 15-metre MSFT and officer safety training (OST) protocols, cross-validating measured heart rate and blood lactate responses achieved between the tests. The initial findings were in agreement with the Roehampton (2003, Unpublished) study finding the level 5:4 standard (equivalent to an aerobic capacity of 35 ml·kg^{-1}·min^{-1}) to be sufficient based on the physiological responses being matched across the protocols (Brewer et al. 2004; Brewer et al. 2013). Following these findings, physiological responses were compared with police training protocols for different specialist roles, enabling the development of aerobic fitness standards for 13 additional roles (Brewer et al. 2010) (as stated in Table 1).

Alternative forms of testing (including the Chester Treadmill Police Walk/Run Test (CTPW/CTPRT) (Sykes, 2016a; Sykes 2016b)) were developed for police officers with injuries/medical conditions that prevented completion of the 15-metre MSFT. The tests were later validated (Morris, Deery, & Sykes, 2019) and accepted by the College of Policing (2014) as suitable alternatives to the 15-metre MSFT.

A secondary objective was to take direct measures (including gas analysis and heart rate) during the 15-metre MSFT. Since the main objective was to investigate the validity and reliability of the CTPW and CTPRT, comparisons between actual and predicted values in the MSFT were not investigated.

Previously the validity and reliability of the 15-metre MSFT has not been investigated using direct gas analysis. Therefore, the study will aim to investigate the aerobic cost of the 15-metre MSFT at 4 target entry levels (5:4, 7:6, 9:4 and 10:5), comparing the estimated aerobic capacity stated in the police fitness standards (Table 1) with direct measures of oxygen uptake, and determine the repeatability (reliability) of this measure. In addition, a sub analyses will be conducted to assess correlation between Respiratory Exchange Ratio (RER) and $\dot{V}O_2$, and gender differences.

Table 1. Police fitness standards for 14 police roles including the Recommended Standard (level:shuttle) and Estimated aerobic capacity ($ml \cdot kg^{-1} \cdot min^{-1}$) adapted from Brewer et al. (2010).

Unit	Recommended standard (level: shuttle)	Estimated aerobic capacity ($ml \cdot kg^{-1} \cdot min^{-1}$)
PST	5: 4	35
Marine police unit	5: 4	35
CBRN	5: 4	35
Method of entry	5: 4	35
Dog handler	5: 7	36
Mounted branch	5: 7	36
Police cyclist	5: 8	36
Police support unit	6: 3	37
Air support	6: 4	37
Police divers	6: 8	39
Marine police (tactical skills)	7: 2	40
Authorized firearms officer	7: 6	41
Armed response vehicle	9: 4	46
Dynamic intervention AFO	10: 5	51

4. Methods

4.1. Participants

Sixty-four police officers (53 male; 11 female) volunteered from police forces based in North-West England and North Wales (age: 41±7 years; height: 177.1±9.2cm; weight: 82.3±13.8kg; body mass index: 26.06±3.2 $kg \cdot m^2$), attending the University of Chester for experimental procedures. Prior to arrival, participants were provided with written information of experimental procedures with additional instruction to abstain from caffeine, alcohol, tobacco and strenuous exercise 24 hours before practice. Before data collection was commenced, participants were required to complete written, informed consent along with health screening procedures (via a health history questionnaire). Additionally, anthropometric measures including stature (wall stadiometer, Seca, Germany) and body mass (Electronic Medical Scale, Seca, Germany) was taken along with Blood pressure and Resting Heart Rate (Omron, Germany).

In accordance with recommendations by Atkinson and Nevill (1998) on validity and reliability trials, the study aimed to recruit a minimum of 40 participants. The selection criteria for taking part in the experiment required participants to be capable of reaching up to level 7:6 based on self-reports. Furthermore, a small number of participants were selected based on being able to achieve up to level 10:5. Participants were not required to run until volitional exhaustion.

4.2. Experimental Design

The experiment was a repeated measures design requiring participants to attend on two separate occasions to complete the 15-metre MSFT, with each test separated by at least 24 hours to enable sufficient recovery from exercise. Examination of repeatability was assessed within two weeks of initial testing for each subject to minimise the effect on time or potential adaptations to fitness.

The method of sampling was based on ability, separating participants into two groups. Group one consisted of participants capable of completing up to level 7:6 without reaching volitional exhaustion, with group two containing participants able to reach level 10:5.

4.3 Measurement procedures

The 15-metre MSFT procedures were conducted in accordance with the College of Policing (2014), reported in the appendices (Appendix E). Briefly, participants repeatedly ran between two cones 15 meters apart with running speed cadence modulated by a pre-recorded audio CD/tape, providing participants with audio cues ("beeps") signifying the beginning/end of a shuttle run. Table 2 illustrates the number of shuttles, speeds and time per shuttle at each level. Before commencement of testing, participants were equipped with portable apparatus for taking continuous measurements of inspired/expired gas (MetaMax 3B, Cortex, Germany) and Heart Rates (Polar s810i, Finland). The 15-metre distance was measured using a standard tape measure. At each end of the 15-metre distance, cones were placed to mark the distance, with a 6-20 ratings of perceived exertion scale (RPE) (Borg, 1982) displayed at both ends of the course.

Participants initially completed a warmup period running the first two levels of the test followed with dynamic stretching. As participants progress through the levels, the time between each bleep is reduced thus gradually increasing both running speed and frequency of shuttles. The test initially commences at 7.88 $km \cdot h^{-1}$, reaching 10.13, 10.94, 11.74 and 12.39 $km \cdot h^{-1}$ at level 5, 7, 9 and 10 respectively.

Throughout the duration of the shuttle running, continuous measurements of heart rate (HR), Oxygen consumption ($\dot{V}O_2$) and RPE were taken. HR was recorded continuously using polar wireless telemetry (Polar s810i, Finland). Recorded HR was applied to two calculations including age predicted maximum heart rate (HRmax) (206-0.7 x age) (Tanaka, Monahan & Seals, 2001) and HR reserve (HRmax – resting HR) (Karvonen, Kentala & Mustala, 1957). $\dot{V}O_2$ and RER were recorded through portable online gas analysis (Metamax 3B, Cortex, Germany) calibrated for breath-by-breath analysis prior to testing. For subsequent analysis, an averaging technique set to 5-second intervals was selected. Robergs, Dwyer, & Astorino, (2010) highlighted there are limitations with data smoothing methods based on time averages, however it was necessary to use 5s averaging to distinguish between shuttles and levels with each shuttle often lasting ~5 seconds each as shown in Table 2.

Table 2. The 15-metre MSFT protocol adapted from Bleep test (2019).

Level	Shuttles	Speed (kph)	Time per shuttles (s)	Total Distance (m)	Total time
1	7	7.88	6.83	105	0:48
2	8	8.52	6.36	225	1:39
3	8	8.69	6.11	345	2:28
4	8	9.33	5.75	465	3:14
5	9	10.13	5.43	600	4:02
6	9	10.62	5.13	735	4:49
7	10	10.94	4.97	885	5:38
8	10	11.42	4.78	1035	6:26
9	10	11.74	4.59	1185	7:11
10	11	12.39	4.39	1350	8:00
11	11	12.87	4.16	1515	8:56
12	12	13.67	3.92	1695	9:33
13	12	14	3.85	1875	10:19
14	13	14.16	3.76	2070	11:08
15	13	14.96	3.59	2265	11:55
16	13	15.28	3.47	2460	12:40
17	14	15.93	3.25	2670	13:35

4.4 Data Analysis

4.4.1. Statistical Analysis

All data was analysed on IBM SPSS version 25 for Windows (IBM Corp, Armonk, NY). All statistical tests were performed in IBM SPSS using an alpha level of $p<0.05$. Normality checks of data was conducted using the Shapiro–Wilk statistic, and descriptive statistics were computed using mean ± standard deviation.

4.4.2. Validity analysis and gender differences

Validity was assessed by investigating the agreement between actual oxygen cost and estimated aerobic capacity using a one sample t-test to examine the mean differences between values. An independent samples t-test was applied to examine gender differences.

4.4.3 Reliability

The test-retest reliability was assessed investigating the agreement between trial 1 (T1) and trial 2 (T2) means for $\dot{V}O_2$ using the 95% Limits of Agreement (LoA) (bias ± 1.96 x SDdiff), Intra-class correlation coefficient (ICC) and Typical error. A paired samples t-test was conducted to investigate if any significant differences were present between test-retest means between trials.

5. Results

5.1 Overview

The physiological responses of trial one for level 5:4, 7:6, 9:4 and 10:5 of the 15-metre MSFT are summarised in Table 3. The physiological responses for trial two are summarised in the appendices (Appendix I, Supplementary Table 1).

Table 3: Physiological Responses at level 5:4; 7:6; 9:4 and 10:5 for Trial One and Comparison to Estimated Aerobic Capacity Required.

Level	**Estimated Aerobic Capacity ($ml.kg^{-1}.min^{-1}$)**	**$\dot{V}O_2$ ($ml.kg^{-1}.min^{-1}$)**	**RER**	**%HRmax**	**%HRR**
5:4	35	40.4±4.8*	1.01±0.09	89±8	82±9
7:6	41	46.8±5.8*	1.07±0.10	94±6	89±7
9:4	46	52.1±7.2*	1.12±0.09	96±5	93±6
10:5	51	55.3±6.5*	1.15±0.09	98±5	95±5

**Significant difference from predicted $\dot{V}O_2$ $p<0.05$.*

5.2. Validity

5.2.1. Level 5:4

A total of 64 (11 female) participants completed T1 up to level 5:4 ($35 ml\cdot kg^{-1}\cdot min^{-1}$), with 49 (nine female) completing up to this level in T2. The one-sample t-test indicated estimated aerobic capacity was significantly lower than measured O_2 cost for both trial 1 and 2 ($40.4±4.8\ ml.kg^{-1}.min^{-1}$ and $40.7±4.7\ ml.kg^{-1}.min^{-1}$, $p<0.05$, respectively), corresponding to a mean difference (MD) of $5.43\ ml\cdot kg^{-1}\cdot min^{-1}$ (95% CI: 4.2 to 6.6) and

5.74 ml·kg^{-1}·min^{-1} (95% CI: 4.4 to 7.1) for T1 and T2, respectively. The mean RPE for this level was 12±2.

5.2.2. Level 7:6

A total of 54 (7 female) participants completed T1 up to level 7:6 (41ml·kg^{-1}·min^{-1}), with 43 (7 female) completing up to this level in T2. The one-sample t-test indicated estimated aerobic capacity was significantly lower than measured O_2 cost for both trial 1 and 2 (46.8±5.8 ml.kg^{-1}.min^{-1} and 46.7±6.3 ml.kg^{-1}.min^{-1}, both $p<0.05$, respectively), with a MD of 5.82 ml·kg^{-1}·min^{-1} (95% CI: 4.2 to 7.4) and 5.67 ml·kg^{-1}·min^{-1} (95% CI: 3.7 to 7.6) for T1 and T2, respectively. The mean RPE for this level was 14±2.

5.2.3. Level 9:4

A total of 31 (2 female) participants completed T1 up to level 9:4 (46 ml·kg^{-1}·min^{-1}), with 23 (2 female) completing up to this level in T2. The one-sample t-test indicated estimated aerobic capacity was significantly lower than measured O_2 cost for both T1 and T2 (52.1±7.2ml.kg^{-1}.min^{-1}, p=.0001 and 51.8±5.2 ml.kg^{-1}.min^{-1}, p=.0001, respectively), with a MD of 6.07 ml·kg^{-1}·min^{-1} (95% CI: 3.4 to 8.7) and 5.75 ml·kg^{-1}·min^{-1} (95% CI: 3.5 to 8.0) for T1 and T2, respectively. The mean RPE for this level was 15±2.

5.2.4. Level 10:5

A total of 19 (0 female) participants completed T1 up to level 10:5 (51ml·kg^{-1}·min^{-1}), with 13 (1 female) completing up to this level in T2. The one-sample t-test indicated estimated aerobic capacity was significantly lower than measured O_2 cost for T1 (55.3±6.5ml.kg-1.min^{-1}, $p<0.05$), however no significant differences were detected in

T2 (54.6±6.3 $ml.kg^{-1}.min^{-1}$, $p>0.05$). This corresponded to a MD of 4.32 $ml \cdot kg^{-1} \cdot min^{-1}$ (95% CI: 1.2 to 7.5) and 3.55 $ml \cdot kg^{-1} \cdot min^{-1}$ (95% CI: -0.3 to 7.4) for T1 and T2 respectively. The mean RPE for this level was 16±2.

5.3. Reliability

The reliability analyses are displayed in Table 4. To conduct the paired samples t-test, only participants that provided two complete sets were included, resulting in a reduction in the sample size at all levels (5:4 (n=43); 7:6 (n=36); 9:4 (n=15) and 10:5 (n=7)). The descriptive statistics for this are included in supplementary Table 2 and 3 (Appendix I). The reliability analyses are displayed in Table 4, including the ninety-five percent Limits of Agreement (95% LoA), ICC and Typical Error. The ICC displayed moderate to excellent reliability between measures (0.65-0.91). Typical error was between 2.76-4.91 $ml \cdot kg^{-1} \cdot min^{-1}$. The paired samples t-test indicated no significant differences between trials for $\dot{V}O_2$ and RER at any of the target levels ($p>0.05$).

Table 4. Reliability analyses of measured $\dot{V}O_2$ between T1 and T2 displaying 95% Limits of Agreement (LoA), Intraclass correlation coefficient (ICC) and Typical Error.

Level: Shuttle	**95% LoA (bias ± 1.96 x SD_{diff})**	**Intraclass Correlation (95% CI)**	**Typical error**
5:4	-0.41±10.2	0.65 (0.35 - 0.81)	±3.69
7:6	-0.57±8.0	0.88 (0.76 - 0.94)	±2.89
9:4	-0.29±7.7	0.91 (0.73 - 0.97)	±2.76
10:5	1.51±13.6	0.72 (-0.80 - 0.95)	±4.91

CI, confidence interval.

5.4. Correlation between RER and $\dot{V}O_2$

The relationship between RER and $\dot{V}O_2$ was determined using a Pearson's correlation coefficient. A significant negative correlation was detected for level 5:4 ($r = -0.463$, $p<0.05$), 9:4 ($r = -0.435$, $p<0.05$) and 10:5 ($r = -0.485$, $p<0.05$), considered a moderate relationship between variables (Cohen, 1988). Upon calculation of the correlation of determination, 21%, 19% and 24% of factors effecting variability are shared by both RER and $\dot{V}O_2$ (Cohen & Holiday, 1996). At level 7:6 a negative relationship was found; however, the correlation was non-significant ($p>0.05$) and the relationship was considered weak ($r=-.226$).

5.5. Gender differences in $\dot{V}O_2$

The measured $\dot{V}O_2$ at each level for each gender are summarised in Table 5. The Independent samples t-tests indicated non-significant differences between genders for $\dot{V}O_2$ at all levels ($p>0.05$) The MD for each level was 1.9 $ml.kg^{-1}.min^{-1}$ (95% CI: -1.3-5.2); 2.0 $ml.kg^{-1}.min^{-1}$ (95% CI: -2.7-6.7) for level 5:4 and 7:6 respectively. Comparisons for level 9:4 and 10:5 were not possible due to an insufficient sample sizes for females.

Table 5. Measured $\dot{V}O_2$ at level 5:4 and 7:6 for males and females.

Gender	**Level 5:4 ($ml \cdot kg^{-1} \cdot min^{-1}$) (95% CI)**	**Level 7:6 ($ml \cdot kg^{-1} \cdot min^{-1}$) (95% CI)**
Female	38.0±4.8 (34.7-41.2)	45.1±3.8 (41.5-48.6)
Male	40.9±4.6 (39.6-42.2)	47.1±6.0 (45.3-48.8)

CI, confidence interval.

6.Discussion

6.1. Validity

The main purpose of the study was to examine and compare the absolute aerobic cost at 4 target entry levels of the 15-metre MSFT (5:4, 7:6, 9:4 & 10:5) against the estimated aerobic capacity determined by Brewer and colleagues (Brewer et al. 2004; Brewer et al. 2010; Brewer et al. 2013) (Table 1). Results of the present study indicated the measured 0_2 cost at these levels was significantly higher by 4-6ml·kg^{-1}·min^{-1} than estimated values required for each of these levels.

Previous research validating the 15-metre MSFT has measured the HR and lactate responses during both the MSFT and specialist training protocols, later used to develop the present standards (Brewer et al. 2004; Brewer et al. 2010; Brewer et al. 2013). The estimated 0_2 values (as presented by Brewer et al. 2010), were developed using a regression equation formulated by the Roehampton study (2003, unpublished). However, a limitation with the majority of these studies was no information of data spread or confidence intervals was reported, thus making comparisons with the present study challenging.

Presently this is the first study to perform gas analysis throughout the 15-metre MSFT, therefore direct comparisons with previous research cannot be made. However, similar to the 4-6 ml.kg^{-1}.min^{-1} difference in the present study, predictive equations/tables for the 20-metre MSFT result in a 6-8 ml.kg^{-1}.min^{-1} lower reported O_2 cost than directly measured values taken during the activity (Direct Method) (Stojanovic et al. 2016; Kilding, Aziz & Teh, 2006; Ruiz et al. 2009; Castagna, Impellizzeri, Manzi, Ditrillo, 2010). Castagna et al. (2010) reported the Ramsbottom, Brewer & Williams (1988) equation significantly underpredicted $\dot{V}O_2max$ by 7.8 $ml.kg^{-1}.min^{-1}$ (Equation: 32.3±3.5 $ml\cdot kg^{-1}\cdot min^{-1}$ vs. Direct method: 40.1±5.0 $ml\cdot kg^{-1}\cdot min^{-1}$, $p<0.001$). Stojanovic et al. (2016) reported the Ramsbottom et al. (1988) equation significantly underpredicted $\dot{V}O_2max$ by 6.5±3.9 $ml\cdot kg^{-1}\cdot min^{-1}$ (Equation: 48.9±4.1 $ml\cdot kg^{-1}\cdot min^{-1}$ vs. Direct Method 55.5±4.9 $ml\cdot kg^{-1}\cdot min^{-1}$, $p<0.001$).

The Ramsbottom et al. (1988) equation as used in the aforementioned studies was developed by using a linear regression equation correlating the number of shuttles completed during the 20-metre MSFT with $\dot{V}O_2max$ attained during a maximal graded treadmill test. Whilst this is concerned with maximal performance opposed to submaximal in the present study, the predictive table /equation used to predict oxygen cost in the 15-metre MSFT was developed using a similar format, correlating the number of shuttles completed with $\dot{V}O_2max$ attained during a method external to the MSFT (i.e. maximal graded treadmill test) (Hazeldine et al. 1995; Roehampton, 2003, Unpublished).

With reports that the energy/oxygen cost of shuttle running is 30-50% higher than straight-line running (as used in the maximal graded treadmill test) (Stevens et al. 2015; Buchheit, Haydar, Hader, Ufland & Ahmaidi, 2011; Buglione & di Prampero, 2013) due to the added accelerations and deaccelerations required (Hatamoto et al. 2013; Hatamoto et al. 2014), it is possible that predictive equations may not account for the additional energy costs required in shuttle running (Flouris, Metsios & Koutedakis, 2005; Ruiz et al. 2009).

There is limited research that suggests redeveloping equations based on directly measured oxygen uptake during the MSFT may improve the accuracy of equations to predict oxygen cost (Silva et al. 2012; Ruiz et al. 2009). Silva et al. (2012) reported non-significant differences in $\dot{V}O_2max$ between an equation based on MSFT performance and portable gas analysis measures during the 20-metre MSFT among Portuguese youth (10-18 years) (Mean difference (MD): 0.0 $ml\cdot kg^{-1}\cdot min^{-1}$; 95% LoA [-11.2-11.2]). In contrast, an equation based on treadmill performance (Léger, Mercier, Gadoury & Lambert, 1988), underpredicted $\dot{V}O_2max$ by -2.9$ml\cdot kg^{-1}\cdot min^{-1}$ (Equation: 45.2±5.9 vs. Measured: 48.1±9.5$ml\cdot kg^{-1}\cdot min^{-1}$) (Silva et al. 2012).

This could suggest redeveloping an equation based on direct measures during the 15-metre MSFT might enhance the accuracy of predictions by negating the energy cost differences between the methods. However research on this topic is limited, mostly using children and adolescents (Silva et al. 2012; Ruiz et al. 2009). Further research is required to affirm these results using adult populations.

The Police fitness standards as outlined by Brewer et al. (2010) suggested officers with a $\dot{V}O_2max$ of 35$ml\cdot kg^{-1}\cdot min^{-1}$ should be able to obtain level 5:4, and a $\dot{V}O_2$ of 41, 46 and 51$ml\cdot kg^{-1}\cdot min^{-1}$ would allow attainment of level 7:6, 9:4 and 10:5 respectively. However, direct measures in the present study indicated O_2 cost was significantly higher at each stage by 4-6$ml\cdot kg^{-1}\cdot min^{-1}$. Upon reaching $\dot{V}O_2max$ during steady state incremental testing, a common occurrence is a $\dot{V}O_2$ plateau, which is an increase in workload despite no further increase in $\dot{V}O_2max$ (Astorino, Rosbergs, Ghiasvand, Marks & Burns, 2000) typically detected in the last 15-30 seconds of maximal exercise testing

(Astorino et al. 2000). During the $\dot{V}O_2$ plateau, the anaerobic metabolism becomes the predominant supply of adenosine triphosphate (ATP) necessary to meet the demands of the activity (Gordon et al. 2011; Astorino et al. 2005).

Although to date this occurrence has not been investigated with the MSFT, it can be suggested upon reaching $\dot{V}O_2$max individuals could continue to exercise for up to 30 seconds longer without a rise in $\dot{V}O_2$, relying on predominantly anaerobic energy sources to service the oxygen deficit being caused. Considering one shuttle can last 3.25-6.83 seconds, this could equate to a few additional shuttles. In relation to the results, a subject with a $\dot{V}O_2$max of 35ml·kg^{-1}·min^{-1} could achieve level 5:4 dependant on the anaerobic metabolism to fulfil the energy required to reach this level.

Upon observation of the data, the mean RER increased from level 5:4 (1.01±0.09) to level 10:5 (1.15±0.09), exceeding 1.00 at all stages. Typically, RER reflects substrate utilisation during steady state exercise up to an RER of 1.00 (Fox, 2016; Widmaier, Raff & Strang, 2006). During non-steady-state, high-intensity exercise, RER will exceed 1.00 as a result of increased CO_2 production and hyperventalation (Deuster & Heled, 2008; Zaggatto, Miyagi, Sakugawa, Kaminagakura & Papoti, 2012). This occurs due to increased buffering of hydrogen and blood lactate produced by working muscles during the anaerobic metabolism (Zaggatto et al. 2012; Issekutz & Rodahl, 1962; Deuster & Heled, 2008). This would suggest the RER above 1.00 in the present study could be indicative of an increased anaerobic contribution to the activity as opposed to achieving $\dot{V}O_2$max (Zagatto et al. 2012).

During an incremental treadmill test, Bertuzzi, Nascimento, Urso, Damasceno, & Lima-Silva (2013) calculated the aerobic-anaerobic contribution at stages leading to volitional exhaustion. At the start of exercise, aerobic contribution was 95.7±1.5%, steadily declining to 86.1±4.7% at $\dot{V}O_2max$ (Bertuzzi et al. 2013). Parallel to this, anaerobic contribution increased from 5.3±1.5% to 13.9±4.7% (Bettuzzi et al. 2013). If applied to the present study, an oxygen uptake of 40 $ml \cdot kg^{-1} \cdot min^{-1}$ found at level 5:4 would constitute 34.4 $ml \cdot kg^{-1} \cdot min^{-1}$ (86%) from aerobic energy sources, whereas the residual 5-6 $ml \cdot kg^{-1} \cdot min^{-1}$ (14%) would represent anaerobic energy sources to service the oxygen deficit. Similarly, an anaerobic contribution of 5-8 $ml \cdot kg^{-1} \cdot min^{-1}$ for level 7:6, 9:4 and 10:5 could be present (Table 6), thereby potentially explaining the differences between estimated and measured values in the present study.

Furthermore, this was supported with a subgroup of participants (n=6) reaching near-exhaustion at level 5:4. With a RER equal to 1.10 (1.10±0.12), a secondary criterion for $\dot{V}O_2max$ (Nelson, Petersen, Dlin, 2010) and predicted maximal HR was reached (100.62±6.04%HRmax), participants recorded a $\dot{V}O_2$ of 37.7±4.5 $ml \cdot kg^{-1} \cdot min^{-1}$. This was lower than the mean value required to meet level 5:4 (40.4±4.7 $ml.kg^{-1}.min^{-1}$), which could suggest anaerobic supplies generated the energy necessary to fulfil the O_2 demands of level 5:4. Interestingly, despite the data suggesting participants had reached near maximal levels, an RPE of 14±1 was recorded suggesting somewhat hard exertion (Borg, 1982).

In contrast, those who reached level 10:5 at the same stage had a $\dot{V}O_2$ similar to the requirements of level 5:4 (41.05±4.95 $ml \cdot kg^{-1} \cdot min^{-1}$) and lower RER of 0.94±0.05

indicating less reliance on anaerobic sources than those achieving up to level 5:4. This was consistent with the negative relationship detected between RER and $\dot{V}O_2$, demonstrating a moderate and significant correlation at level 5:4, 9:4 and 10:5 ($r=-0.435-0.485$, $p<0.05$). This suggested the lower reported $\dot{V}O_2$ values at each level were associated with higher RER values and vice versa.

Based on these findings, officers achieving a $\dot{V}O_2max$ of 35ml·kg^{-1}·min^{-1} should theoretically be able to achieve the minimal standard (level 5:4) of the 15-metre MSFT, thereby supporting the previous findings of Brewer and colleagues (Brewer et al. 2010; Brewer et al. 2013). Further investigations should focus on participants with a $\dot{V}O_2max$ of 35ml·kg^{-1}·min^{-1} examining the probability of achieving level 5:4 to affirm the current findings.

Table 6. The estimated aerobic and anaerobic energy contribution at $\dot{V}O_2max$ upon level 5:4, 7:6, 9:4 and 10:5 of the 15-metre MSFT, based on the findings of Bertuzzi et al. (2013).

Level and measured oxygen cost.	Aerobic contribution (86%) in ml·kg^{-1}·min^{-1}	Anaerobic contribution (14%) in ml·kg^{-1}·min^{-1}.
5:4 (40 ml·kg^{-1}·min^{-1})	34.4	5.6
7:6 (46.8 ml·kg^{-1}·min^{-1})	40.42	6.4
9:4 (52 ml·kg^{-1}·min^{-1})	44.8	7.2
10:5 (55 ml·kg^{-1}·min^{-1})	47.5	7.5

6.2. Reliability

The reliability analysis indicated non-significant bias between trials ($p>0.05$), with moderate to high correlations between trials (ICC=0.65-0.91). The LoA analysis indicated random error was at best 7.7ml·kg^{-1}·min^{-1} with a systematic bias of -0.29ml·kg^{-1}·min^{-1} (at level 9:4). Based on the results from T1 (52.1±7.2 ml.kg^{-1}.min^{-1}), participants would be expected to obtain a $\dot{V}O_2$ between 44.11-59.51 ml.kg^{-1}.min^{-1} upon test repetition. The worst random error detected was 13.6 ml.kg^{-1}.min^{-1} and a systematic bias of 1.51 ml.kg^{-1}.min^{-1} (level 10:5). When applied to T1 results (55.3±6.5 ml.kg^{-1}.min^{-1}), a retest would expect a $\dot{V}O_2$ between 43.21-67.39 ml.kg^{-1}.min^{-1}. It would be estimated that random error for each level would represent 25.25%, 17.1%, 14.78% and 24.6% of the variance for level 5:4, 7:6, 9:4 and 10:5, respectively.

Previous studies investigating reliability of the 20-metre MSFT (Cooper, Baker, Tong, Roberts & Hanford, 2005; Aandstad, Holme, Berntsen & Anderssen, 2011, Lamb & Rogers, 2007) has reported ±95% LoA of -0.4±2.7 ml·kg^{-1}·min^{-1} (Cooper et al. 2004), −1.1±4.7 ml·kg^{-1} ·min^{-1} (Lamb & Rogers, 2007) and -0.8±3.1 ml·kg^{-1}·min^{-1} (Aandstad et al. 2011) all reporting non-significant differences between trials ($p>0.05$). However, the aforementioned studies examined the reliability of predicted $\dot{V}O_2$max, whilst the present study is the first to examine the reliability of directly measured O_2.

With interpreting the results from the present study, it has been recommended to consider if test-retest reliability would mask any changes in $\dot{V}O_2$max that might be expected from interventions or training programmes (Morris, Lamb, Hayton, Cotterrell, Buckley, 2010; Morris, Lamb, Cotterell & Buckley, 2009). A typical 6-12-week aerobic fitness programme would expect $\dot{V}O_2max$ increases between 10-20% (Metcalfe, Babraj, Fawkner & Vollaard, 2012; Hiruntrakul, Nanagara, Emasithi & Borer, 2010; Matsuo et al. 2014) with a baseline fitness between 33-42 $ml.kg^{-1}\cdot min^{-1}$, similar to populations reaching level 5:4 and 7:6 (40.4±4.8; 46.8±5.8 $ml\cdot kg^{-1}\cdot min^{-1}$, respectively). With random error between 17.1-25.5% at these levels for 95% LoA, the MSFT would likely mask any changes in $\dot{V}O_2max$ upon retest. However, some authors have suggested the 95% LoA can be too stringent, instead advocating the use of typical error (standard error of estimate) (Morris et al. 2010; Morris et al. 2009; Hopkins, 2000). Typical error would represent <10% random error with 9.1% error (±3.69 $ml\cdot kg^{-1}\cdot min^{-1}$) between T1 and T2 for level 5:4 and 6.2% error (±2.89 $mL\cdot kg^{-1}\cdot min^{-1}$) for level 7:6. This would suggest reliability was acceptable with improvements expected of >10%, however it should be noted typical error only represents two thirds (68%) of the sample opposed to LoA covering 95% (Atkinson & Nevill, 2000).

Amongst populations with a high starting baseline $\dot{V}O_2max$ (48-60 $ml\cdot kg^{-1}\cdot min^{-1}$), similar to participants reaching level 9:4 and 10:5 (52.1±7.2 and 55.3±6.5 $ml\cdot kg^{-1}\cdot min^{-1}$ respectively), less pronounced improvements of 4-10% from an aerobic fitness programme would be expected (Helgerud et al. 2007; Helgerud, Engen, Wilsloff & Hoff, 2001; Williams, Rayson & Jones, 1999; Rowan, Kueffner, Stavrianeas, 2012). With 95% LoA representing 14.78-24.6% of random error and typical error being 5.1% (±2.76 $ml\cdot kg^{-1}\cdot min^{-1}$) for level 9:4 and 8.9% (±2.89 $ml\cdot kg^{-1}\cdot min^{-1}$) for level 10:5, this would suggest improvement in $\dot{V}O_2max$ could be masked upon retest for these populations. Thus, the test would be unreliable for populations with high baseline $\dot{V}O_2max$.

6.3. Gender differences

Previous research has highlighted gender differences in measurement of $\dot{V}O_2max$, often with females producing significantly lower mean outputs (Loe, Rognmo, Saltin & Wisløff, 2013; Sandbakk, Ettema, Lierdal & Holmberg, 2012). Loe et al. (2013) reported among men (n=1929) and women (n=1881) of various fitness levels aged 20-90 years, the mean $\dot{V}O_2max$ of women was significantly lower by 18.7% (37.0±7.5 $ml \cdot kg^{-1} \cdot min^{-1}$ vs. 45.4±8.9 $ml \cdot kg^{-1} \cdot min^{-1}$, $p<0.001$). Whilst females in the present study had lower $\dot{V}O_2$ at all levels (table 5), the independent samples t-test indicated small non-significant mean differences between 1.9 to 2.0 $ml.kg^{-1}.min^{-1}$ ($p>0.05$).

However, the previous studies investigating gender differences (such as Loe et al. 2013) examined maximal performance opposed to sub-maximal in the present study, thus potentially explaining differences. In agreement with the present findings, Maldondo-Martin, Mujika & Padilla (2004) reported whilst $\dot{V}O_2max$ was significantly greater amongst male middle distance and marathon runners opposed to females during a maximal treadmill protocol (69.3±3.5 vs. 60.7±4.7 $ml \cdot kg^{-1} \cdot min^{-1}$, respectively, $p<0.01$), at matched velocities of 4.03-4.87 $m \cdot s^{-1}$, oxygen uptake was not significantly different ($p>0.05$) (4.03 $m \cdot s^{-1}$: 45.6±3.8 vs. 44.8±3.9 $ml \cdot kg^{-1} \cdot min^{-1}$; 4.45 $m \cdot s^{-1}$: 51.2±3.3 vs. 50.7±3.6 $ml \cdot kg^{-1} \cdot min^{-1}$; 4.87 $m \cdot s^{-1}$: 56.5±3.2 vs. 55.32±4.3 $ml \cdot kg^{-1} \cdot min^{-1}$, respectively). Therefore these findings are consistent with the present study suggesting gender differences in oxygen cost are negated when velocity/exercise intensity is matched.

6.4. Strengths, Limitations and Future Recommendations

Due to this investigation consisting of serving UK police officers, the findings provide relevancy and generalisability to the population. Moreover, the inclusion of both male and female participants further adds to the generalisation of results, which can often be a shortcoming of validity and reliability trials (Cooper et al. 2005; Aandstad et al. 2011; Flouris et al. 2004).

Whilst there are noted errors with commercially available metabolic analysers (Hodges, Brodie, Bromley, 2005), the device used within the present study (MM3B) is found to be stable and reliable with no greater than 3% technical error of measurement for $\dot{V}O_2$, $\dot{V}CO_2$ and $\dot{V}E$ (Mcfarlane & Wong, 2012; Polese et al. 2015). Secondly, the same device was used throughout all trials to control for potential variation between devices.

Despite attempts made to include officers from Black, Asian and minority ethnic (BAME) groups, the current sample consisted of only white males and females, weakening the generalisation of the findings. Difficulty in recruitment is likely due to low employment of BAME making up only 6.6% of the UK police force (based on 2018 figures) (https://www.ethnicity-facts-figures.service.gov.uk/workforce-and-business/workforce-diversity/police-workforce/latest#download-the-data). Some observational studies have suggested CRF differences can occur among ethnicities (Swift et al. 2017; Sanders & Duncan, 2006; Duncan, Li & Zhou, 2005), therefore a repeat of this study focusing on ethnic minorities maybe warranted to see if the present results are upheld. Not all police officers were able to attend a repeat trial due to work constraints, or individuals improving performance on the second visit, therefore the sample size for the reliability analysis was reduced from the validity trial. This was particularly prominent at level 7:6 (n=36), level 9:4 (n=15) and 10:5 (n=7), falling below the recommendations outlined by Atkinson & Nevil (1998) for a minimum sample size of 40 for reliability analyses, thereby potentially weakening the analysis. Prior to testing, directly measured $\dot{V}O_2$max of individuals was not established. This therefore weakens the analysis as direct comparisons could not be made with the present data, instead relying on secondary criterions of $\dot{V}O_2$max (i.e. HR and RER) to assume attainment of $\dot{V}O_2$max.

7.Conclusion

Data from the present study demonstrates that absolute measurements of oxygen uptake at level 5:4; 7:6; 9:4 and 10:5 of the 15-metre MSFT is significantly higher (4-6 ml) than the estimated values published by Brewer et al. (2010). The differences between the estimated aerobic capacity required and actual O_2 cost are potentially explained by an anaerobic contribution near exhaustion. Further research should focus on using participants with a $\dot{V}O_2$max approximately 5 ml·kg^{-1}·min^{-1} below the requirements for each target stage to test this theory.

Dependant on the $\dot{V}O_2$max level of the subject, the reliability of the test would be considered acceptable for low-fit populations with a less than 10% error found between test-retest measures. However, this would likely be less reliable among high-fit individuals where improvement in oxygen uptake from exercise conditioning would be less detectable.

8. References

Aandstad, A., Holme, I., Berntsen, S., & Anderssen, S. A. (2011). Validity and reliability of the 20-meter shuttle run test in military personnel. *Military Medicine, 176*(5), 513-518. doi:10.7205/MILMED-D-10-00373.

Anderson, G. S., Plecas, D., & Segger, T. (2001). Police officer physical ability testing - re-validating a selection criterion. *Policing: An International Journal of Police Strategies & Management, 24*(1), 8-31. doi:10.1108/13639510110382232.

Astorino, T. A., Willey, J., Kinnahan, J., Larsson, S. M., Welch, H., & Dalleck, L. C. (2005). Elucidating determinants of the plateau in oxygen consumption at VO_{2max}. *British Journal of Sports Medicine, 39*(9), 655-660. doi:10.1136/bjsm.2004.016550.

Astorino, T.A., Robergs, R.A., Ghiasvand, F., Marks, D., & Burns, S. (2000) Incidence of the oxygen plateau at VO2max during exercise testing to volitional fatigue. *Journal of Exercise Physiology Online*, *3*(4), 1-12.

Atkinson, G., Nevill, A, M. (1998) Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Medicine*. *26*(4), 217-38. doi:10.2165/00007256-199826040-00002.

Bertuzzi, R., Nascimento, E. M. F., Urso, R. P., Damasceno, M., & Lima-Silva, A. E. (2013). Energy system contributions during incremental exercise test. *Journal of Sports Science & Medicine, 12*(3), 454-460.

Bleep Test (2019). *Bleep Test Table (15m): by www.bleeptest.co.uk.* Retrieved September 09, 2019 from http://www.bleeptest.co.uk/wp-content/uploads/2017/03/15m-Bleep-Test-Table.pdf.

Bonneau, J., Brown, J. (1995) Physical ability, fitness and police work. *Journal of Clinical Forensic Medicine.* 2, 157-164.

Borg, G. A. (1982). Psychophysical bases of perceived exertion. *Medicine and Science in Sports and Exercise, 14*(5), 377-381.

Brewer, J., Breivik, S., Eccleston, S., Howells, S., Morgan, N., Bezodis, N., & Fairclough, S. (2004). *Lilleshall Report: Research into Fitness for the Police Service*. Retrieved from UK: Lilleshall Sports Injury and Human Performance Centre.

Brewer, J., Buckle, P., Castle, P. (2013). The use of workplace physiological measurements to establish the minimum fitness standards required for entry into the United Kingdom police service. *Journal of Athletic Enhancement*, 2(4) doi:10.4172/2324-9080.1000121.

Brewer, J., Morgan, N., Breivik, S., Dyer, A & Harmer, D. (2010). *Job Related Fitness Tests for Police Officer Specialist Posts*. UK: National Police Improvement Agency. Retrieved from the College of policing website: https://www.college.police.uk/What-we-do/Standards/Fitness/Documents/FWG_Report_final.pdf.

Buchheit, M., Haydar, B., Hader, K., Ufland, P., & Ahmaidi, S. (2011). Assessing running economy during field running with changes of direction: Application to 20 m shuttle runs. *International Journal of Sports Physiology and Performance, 6*(3), 380-395. doi:10.1123/ijspp.6.3.380.

Buglione, A., & di Prampero, P. E. (2013). The energy cost of shuttle running. *European Journal of Applied Physiology, 113*(6), 1535-1543. doi:10.1007/s00421-012-2580-2589.

Castagna, C., Impellizzeri, F. M., Chamari, K., Carlomagno, D., & Rampinini, E. (2006). Aerobic fitness and yo-yo continuous and intermittent tests performances in soccer players: A correlation study. *Journal of Strength and Conditioning Research*, 20(2), 320-325. doi:10.1519/R-18065.1.

Castagna, C., Impellizzeri, F. M., Manzi, V., & Ditroilo, M. (2010). The assessment of maximal aerobic power with the multistage fitness test in young women soccer players. *Journal of Strength and Conditioning Research, 24*(6), 1488-1494. doi:10.1519/JSC.0b013e3181d8e97a.

Cohen, L., & Holiday, M.G. (1996). *Practical statistics for students: An introductory test.* London: Paul Chapman Publishing.

Cohen. J. (1988). *Statistical power analysis for the behavioural sciences.* (2^{nd} ed.) Hillsdale, NJ: Erlbaum.

College of Policing (2014) *Implementation. Job-Related Fitness Tests for the Police Service of England and Wales*: Winsor Recommendation 33. Retrieved from https://www.college.police.uk/What-we-do/Standards/Fitness/Documents/PoliceFitnessGuidance.pdf.

Cooper, S., Baker, J. S., Tong, R. J., Roberts, E., & Hanford, M. (2005). The repeatability and criterion related validity of the 20 m multistage fitness test as a predictor of maximal oxygen uptake in active young men. *British Journal of Sports Medicine*, *39*(4), 1-7. doi:10.1136/bjsm.2004.013078.

Deuster, P & Heled, Y. (2008). Testing for Maximal Aerobic Power. *The Sports Medicine Resource Manual*. 520-528. 10.1016/B978-141603197-0.10069-2.

Duncan, G. E., Li, S. M., & Zhou, X. (2005). Cardiovascular fitness among U.S. adults: NHANES 1999-2000 and 2001-2002. *Medicine and Science in Sports and Exercise*, 37(8), 1324-1328. doi:10.1249/01.mss.0000174893.74326.11.

Flouris, A. D., Koutedakis, Y., Nevill, A., Metsios, G. S., Tsiotra, G., & Parasiris, Y. (2004). Enhancing specificity in proxy-design for the assessment of bioenergetics. *Journal of Science and Medicine in Sport, 7*(2), 197-204. doi:10.1016/S1440-2440(04)80009-4.

Flouris, A. D., Metsios, G. S., & Koutedakis, Y. (2005). Enhancing the efficacy of the 20 m multistage shuttle run test. *British Journal of Sports Medicine, 39*(3), 166-170. doi:10.1136/bjsm.2004.012500.

Fox, S. (2016). *Human physiology*. (14^{th} edition). New York, NY: McGraw-Hill.

Gordon, D., Hopkins, S., King, C., Keiller, D., & Barnes, R. J. (2011). Incidence of the plateau at VO_{2max} is dependent on the anaerobic capacity. *International Journal of Sports Medicine, 32*(1), 1-6. doi:10.1055/s-0030-1267192.

Hatamoto, Y., Yamada, Y., Fujii, T., Higaki, Y., Kiyonaga, A., & Tanaka, H. (2013). A novel method for calculating the energy cost of turning during running. *Journal of Sports Medicine, 9*(4), 117-122. doi:10.2147/OAJSM.S39206.

Hatamoto, Y., Yamada, Y., Sagayama, H., Higaki, Y., Kiyonaga, A., & Tanaka, H. (2014). The relationship between running velocity and the energy cost of turning during running. *PloS One*, *9*(1), e81850-e81850. doi:10.1371/journal.pone.0081850.

Hazeldine, R., Lakomy, H, K, A., & Simpson, D. (1995) *Recommendations for the Standards of Fitness of Metropolitan Police Officers*. Unpublished research, Department of Physical Education, Sports Science and Recreation Management, Loughborough University of Technology, UK.

Helgerud, J., Engen, L. C., Wisloff, U., & Hoff, J. (2001). Aerobic endurance training improves soccer performance. *Medicine and Science in Sports and Exercise, 33*(11), 1925-1931. doi:10.1097/00005768-200111000-00019.

Helgerud, J., Høydal, K., Wang, E., Karlsen, T., Berg, P., Bjerkaas, M., . . . Hoff, J. (2007). Aerobic high-intensity intervals improve VO2max more than moderate training. *Medicine and Science in Sports and Exercise, 39*(4), 665-671.

Hiruntrakul, A., Nanagara, R., Emasithi, A., & Borer, K. T. (2010). Effect of once a week endurance exercise on fitness status in sedentary participants. *Journal of the Medical Association of Thailand, 93*(9), 1070-4.

Hodges, L. D., Brodie, D. A., & Bromley, P. D. (2005). Validity and reliability of selected commercially available metabolic analyzer systems. *Scandinavian Journal of Medicine & Science in Sports, 15*(5), 271-279. doi:10.1111/j.1600-0838.2005.00477.x.

Hopkins, W.G. (2000). Measures of reliability in sports medicine and science. *Sports Medicine, 30*(1), 1-15. doi:10.2165/00007256-200030010-00001.

Issekutz, J., Birkhead, N. C., & Rodahl, K. (1962). Use of respiratory quotients in assessment of aerobic work capacity. *Journal of Applied Physiology, 17*(1), 47-50. doi:10.1152/jappl.1962.17.1.47.

Karvonen, M.J., Kentala, E. & Mustala, O. (1957). The effects of training on heart rate; a longitudinal study. *Annales Medicinae Experimentalis et Biologiae Fenniae*, *35*(3): 307-315.

Kilding, A. E., Aziz, A. R., & Teh, K. C. (2006). Measuring and predicting maximal aerobic power in international-level intermittent sport athletes. *The Journal of Sports Medicine and Physical Fitness*, *46*(3), 366-372.

Kyröläinen, H., Häkkinen, K., Kautiainen, H., Santtila, M., Pihlainen, K., & Häkkinen, A. (2008). Physical fitness, BMI and sickness absence in male military personnel. *Occupational Medicine (Oxford, England)*, 58(4), 251-256.

Lamb, K. L., & Rogers, L. (2007). A re-appraisal of the reliability of the 20 m multi-stage shuttle run test. *European Journal of Applied Physiology, 100*(3), 287-292. doi:10.1007/s00421-007-0432-9.

Léger, L. A., & Lambert, J. (1982). A maximal multistage 20-m shuttle run test to predict VO2 max. *European Journal of Applied Physiology and Occupational Physiology, 49*(1), 1-12.

Léger, L. A., Mercier, D., Gadoury, C., & Lambert, J. (1988). The multistage 20 metre shuttle run test for aerobic fitness. *Journal of Sports Sciences, 6*(2), 93-101.

Loe, H., Rognmo, Ø., Saltin, B., & Wisløff, U. (2013). Aerobic capacity reference data in 3816 healthy men and women 20-90 years. *PloS One, 8*(5), e64319-e64319. doi:10.1371/journal.pone.0064319.

Macfarlane, D. J., & Wong, P. (2012). Validity, reliability and stability of the portable cortex metamax 3B gas analysis system. *European Journal of Applied Physiology, 112*(7), 2539-2547. doi:10.1007/s00421-011-2230-7.

Maldonado-Martin, S., Mujika, I., & Padilla, S. (2004). Physiological variables to use in the gender comparison in highly trained runners. *The Journal of Sports Medicine and Physical Fitness, 44*(1), 8-14.

Matsuo, T., Saotome, K., Seino, S., Eto, M., Shimojo, N., Matsushita, A., . . . Mukai, C. (2014). Low-volume, high-intensity, aerobic interval exercise for sedentary adults: VO2max, cardiac mass, and heart rate recovery. *European Journal of Applied Physiology, 114*(9), 1963-1972. doi:10.1007/s00421-014-2917-7.

Metcalfe, R. S., Babraj, J. A., Fawkner, S. G., & Vollaard, N. B. J. (2012). Towards the minimal amount of exercise for improving metabolic health: Beneficial effects of reduced-exertion high-intensity interval training. *European Journal of Applied Physiology,* 112(7), 2767-2775. doi:10.1007/s00421-011-2254-z.

Morris, M., Deery, E., Sykes, K. (2019). Chester Treadmill police test as alternatives to 15-m shuttle running. *Occupational medicine (London)*, *69* (2), 133-138. doi: 10.1093/occmed/kqz014.

Morris, M., Lamb, K. L., Hayton, J., Cotterrell, D., & Buckley, J. (2010). The validity and reliability of predicting maximal oxygen uptake from a treadmill-based sub-maximal perceptually regulated exercise test. *European Journal of Applied Physiology, 109*(5), 983-988. doi:10.1007/s00421-010-1439-1.

Morris, M., Lamb, K., Cotterrell, D., & Buckley, J. (2009). Predicting maximal oxygen uptake via a perceptually regulated exercise test (PRET). *Journal of Exercise Science & Fitness, 7*(2), 122-128. doi:10.1016/S1728-869X(09)60015-0.

Nelson, M. D., Petersen, S. R., & Dlin, R. A. (2010). Effects of age and counselling on the cardiorespiratory response to graded exercise. *Medicine and Science in Sports and Exercise*, 42(2), 255-264. doi:10.1249/MSS.0b013e3181b0e534.

Pate, R. R., & Kriska, A. (1984). Physiological basis of the sex difference in cardiorespiratory endurance. *Sports Medicine, 1*(2), 87-98.

Polese, J.C., Ada, L., Parreira, V.F., Faria, G.S., Avelino, P., Teixeira-Salmela, L.F. (2015). Test-retest Reliability of Cardiorespiratory Variables Measured with the Metamax 3b During the Six-minute Walking Test after Stroke. *Physical Medicine and Rehabilitation International,* 2(1): 1028-1032.

Ramsbottom, R., Brewer, J., & Williams, C. (1988). A progressive shuttle run test to estimate maximal oxygen uptake. *British Journal of Sports Medicine, 22*(4), 141-144. doi:10.1136/bjsm.22.4.141.

Robergs, R. A., Dwyer, D., & Astorino, T. (2010). Recommendations for improved data processing from expired gas analysis indirect calorimetry. *Sports Medicine, 40*(2), 95-111.

Roehampton University of Surrey. (2003, unpublished study). *Report on the Validation of the 15m Work Related Fitness Test*. UK: Roehampton University of Surrey.

Rowan, A.E., Kueffner, T.E & Stavrianeas, S. (2012). Short Duration High-Intensity Interval Training Improves Aerobic Conditioning of Female College Soccer Players. *International Journal of Exercise Science*, 5(3), 232-238**.**

Ruiz, J. R., Silva, G., Oliveira, N., Ribeiro, J. C., Oliveira, J. F., & Mota, J. (2009). Criterion-related validity of the 20-m shuttle run test in youths aged 13-19 years. *Journal of Sports Sciences, 27*(9), 899-906. doi:10.1080/02640410902902835.

Sandbakk, Ø., Ettema, G., Leirdal, S., Holmberg, H. (2012). Gender differences in the physiological responses and kinematic behaviour of elite sprint cross-country skiers. *European Journal of Applied Physiology, 112*(3), 1087-1094. doi:10.1007/s00421-011-2063-4.

Sanders, L. F., & Duncan, G. E. (2006). Population-based reference standards for cardiovascular fitness among U.S. adults: NHANES 1999-2000 and 2001-2002.

Medicine and Science in Sports and Exercise, 38(4), 701-707. doi:10.1249/01.mss.0000210193.49210.b5.

Silva, G., Oliveira, N. L., Aires, L., Mota, J., Oliveira, J., & Ribeiro, J. C. (2012). Calculation and validation of models for estimating VO_{2max} from the 20-m shuttle run test in children and adolescents. *Archives of Exercise in Health and Disease, 3*, 145-152. doi:10.5628/aehd.v3i1-2.20.

Stevens, T. G. A., De Ruiter, C. J., Van Maurik, D., Van Lierop, C. J. W., Savelsbergh, G. J. P., & Beek, P. J. (2015). Measured and estimated energy cost of constant and shuttle running in soccer players. *Medicine and Science in Sports and Exercise*, *47*(6), 1219-1224. doi:10.1249/MSS.0000000000000515.

Stojanovic, M. D., Calleja-Gonzalez, J., Mikic, M., Madic, D. M., Drid, P., Vuckovic, I., & Ostojic, S. M. (2016). Accuracy and criterion-related validity of the 20-M shuttle run test in well-trained young basketball players. *Montenegrin Journal of Sports Science and Medicine*, *5*(2), 5-10.

Strating, M., Bakker, R. H., Dijkstra, G. J., Lemmink, K A P M, & Groothoff, J. W. (2010). A job-related fitness test for the dutch police. *Occupational Medicine (Oxford, England), 60*(4), 255-260.

Swift, D.L., Nevels, T.R., Landrine, H., Lavie, C.J., Newton, R.L., Das, B.M., Vos, P., Houmard, J.A. (2017). Abstract 32: Racial Differences in Cardiorespiratory Fitness Between African and Caucasian Americans: A Meta-analysis. *American Heart Association*, *135* (32), A32.

Sykes K. (2016a) *The Chester Treadmill Police Walk Test.* In: Sykes K: *The Chester Aerobic Fitness Tests Manual* (pp. 27-31). Chester, UK: Cartwright Fitness Publications.

Sykes K. (2016b) *The Chester Treadmill Police Run Test*. In: Sykes K: *The Chester Aerobic Fitness Tests Manual* (pp. 32-35). Chester, UK: Cartwright Fitness Publications.

Tanaka, H., Monahan, K, D., Seals, D, R. (2001) Age-predicted maximal heart rate revisited. *Journal of the American College of Cardiology*, *37*(1), 153-6.

Widmaier, E., Raff, H & Strang, K. (2006). *Vanders Human Physiology*. (10th edition). New York, NY: McGraw-Hill.

Williams, A. G., Rayson, M. P., & Jones, D. A. (1999). Effects of basic training on material handling ability and physical fitness of British army recruits. *Ergonomics, 42*(8), 1114-1124. doi:10.1080/001401399185171.

Winsor, T. (2013). *Winsor review*. United Kingdom: Gov.uk.

Zagatto, A. M., Miyagi, W. E., Sakugawa, R. L., Kaminagakura, E. I., & Papoti, M. (2012). Aerobic endurance measurement by respiratory exchange ratio during a cycle ergometer graded exercise test. *Journal of Exercise Physiology Online, 15*(5), 49-56.

9. Appendices

Appendix A

PARTICIPANT INFORMATION SHEET

Validity and reliability of the 15m Multi-Stage Fitness Test, Chester Treadmill Walk Test (Police) and Chester Treadmill Test (Firearms) in predicting fitness levels ($\dot{V}O_2$) required for specific Police Force Units

You are being invited to take part in a research study. Before you decide, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully. If you would like further information, please do not hesitate to ask our research team.

What is the purpose of the study?

The aim of the study is to investigate the validity and reliability of the 15 metre multi-stage fitness test (15m MSFT) currently used in Police fitness testing and two proposed alternatives, Chester treadmill walking test (CTWT) and Chester Treadmill Test (CTT), as job-related fitness tests for police officers

Why have I been chosen?

You have been chosen due to you being the appropriate age, being apparently healthy and being a serving Police officer.

Do I have to take part?

No, all potential participants are volunteers and you should not feel obliged or prejudiced in any way either by the researchers, your employer, or your work colleagues to participate. If you do agree to participate you are free to withdraw at any time without reason and without prejudice to your work. Your employer will not be made aware of reason for withdrawal

What will be required of me if I take part?

If you wish to participate, you will be asked to attend the Exercise Physiology laboratory and the Small Hall at the University of Chester, Parkgate Road, Chester on two occasions. You will be asked to abstain from heavy exercise, alcohol, tobacco and caffeine for 24 hours prior to testing.
On each visit you will be asked to participate in one 15m MSFT and one treadmill test, which will be performed wearing a face mask (to measure the air you breathe) and chest strap heart

rate monitor. You will repeat the same exercise tests on your second visit to the laboratory. Measures including blood pressure, body mass and height will be taken prior to testing. A health screening questionnaire must be completed along with an informed consent form prior to beginning the study.

During the exercise testing we will be measuring the following:

1. Oxygen consumption and carbon dioxide production (for which you will be required to wear a face mask).
2. Heart rate (for which you will be required to wear a belt around your chest).
3. Rating of perceived exertion (a measurement of how hard it feels while you are exercising).

What are the possible disadvantages and risks of taking part?

Risk assessments are carried out by researchers prior to carrying out research to help manage risks associated with this study. You may feel slight discomfort during exercise testing due to exercise intensity and/or online gas analysis face mask or vest worn during testing. As with any exercise or shuttle running there is a risk of injury due to slips, trips or falls but these will be minimised by ensuring that the flooring is suitable for shuttle running and there are no electrical leads etc., which may cause trips.

What are the possible benefits of taking part?

By taking part, you will be contributing to the development and understanding of the current fitness testing procedure used by the Police. You will be contributing to the development of an alternative walking treadmill tests which may be used in place of the 15m MSFT. Participating will also allow you valuable practice of the exercise testing used within the Police force.

What will happen to the results of the research study?

The results will be written up in a report as part an MSc thesis and also possibly used for research publication. A report of sample data and relevant findings will also be given to the College of Policing. Individuals who participate will not be identified in any subsequent report or publication.

What if something goes wrong?

If you wish to complain or have any concerns about any aspect of the way you have been approached or treated during the course of this study, please contact Dean of the Faculty of Life Sciences, University of Chester, Parkgate Road, Chester, CH1 4BJ, 01244 513055.

Will my taking part in the study be kept confidential?

All information which is collected during the course of the research will be kept strictly confidential so that only the researcher/s carrying out the data collection will have access to such information.

Who may I contact for further information?

Many thanks for your due consideration, if there is anything that is not clear or if you would like more information, please do not hesitate to ask our research team.

- Dr Mike Morris (University of Chester) m.morris@chester.ac.uk 01244513363
- Miss Elizabeth Parker (University of Chester) e.parker@chester.ac.uk 01244513959

Appendix B.

University of Chester

<u>Informed Consent</u>

Validity and reliability of the 15m Multi-Stage Fitness Test, Chester Treadmill Walk Test (Police) and Chester Treadmill Test (Firearms) in predicting fitness levels ($\dot{V}O_2max$) required for specific Police Force Units

Please initial box

I confirm that I have read and understand the information sheet for the above study and have had the opportunity to ask questions. ☐

I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason and without my legal rights being affected. ☐

I agree to take part in the above study. ☐

Name of Participant	Date	Signature
____________________	__________________	______________
Researcher	Date	Signature
____________________	__________________	______________

Appendix C.

Pre-exercise test questionnaire

Participant Name: _____________ **Date of birth:** __________________
BP: ____________ **Height**: ____________________ **Weight**:

In order to ensure that this study is as safe and accurate as possible, it is important that each potential participant is screened for any factors that may influence the study.

Please circle your answer to the following questions:

1. Has your doctor ever said that you have a heart condition and that you should only perform physical activity recommended by a doctor? YES / NO
2. Do you feel pain in the chest when you perform physical activity? YES / NO
3. In the past month, have you had chest pain when you were not performing physical activity? YES / NO
4. Do you lose your balance because of dizziness or do you ever lose consciousness? YES / NO
5. Do you have bone or joint problems (e.g. back, knee or hip) that could be made worse by a change in your physical activity? YES / NO
6. Is your doctor currently prescribing drugs for your blood pressure or heart condition? YES / NO
7. Are you pregnant, or have you been pregnant in the last six months? YES / NO
8. Have you injured your hip, knee or ankle joint in the last six months? YES / NO
9. Do you know of any other reason why you should not participate in physical activity? YES / NO

Thank you for taking your time to fill in this form. If you have answered 'yes' to any of the above questions, please speak to a member of our research team.

Signed (Participant): ____ Date____________

Signed (Researcher):____

Appendix D.

Inclusion/exclusion criteria

Inclusion Criteria

- Police officers currently signed off by Occupational Health as fit for service within one of the following units: Marine Police Unit, CBRN, Method of Entry, Dog Handler, Mounted Branch, Police Cyclist, Police Support Unit, Air Support, Police Divers, Marine Police (Tactical Skills), Authorised Firearms Officer, Armed Response Vehicle, Dynamic Intervention AFO
- Males and females aged 18-65
- Able to attend the exercise physiology laboratory to carry out 15m MSFT and relevant treadmill test on two occasions
- Satisfied health screening questionnaire

Exclusion Criteria

- Those not fulfilling the above inclusion criteria
- Anyone with a current acute or chronic neuromusculoskeletal or circulatory condition, which may be worsened by included exercise tests
- Anyone who has chronic obstructive pulmonary disease
- Anyone who at the time of the study has or develops an acute bout of bronchitis or asthma, fever/flu, infection, or viral condition causing mild muscle fatigue, or any acute lung disorder where breathing responses to mild to moderate levels of exertion result in abnormal breathing
- Anyone having suffered an acute cardiac event within the last 6 months, who has not yet been fully stabilised and discharged from the care of a cardiologist or specialist physician and unable to easily perform mild to moderate physical activities
- Anyone who knows she is pregnant
- Anyone who has any other health condition (mental or physical) in which they feel would be exacerbated by the stated exercise tests

Appendix E.

15-metre MSFT protocol

Equipment: flat, non-slip surface of more than 15 meters, RPE scale, marker cones, 15-meter tape measure, beep test audio CD and CD player, Cortex Metamax 3B, heart rate monitor.

• A 15-meter distance will be measured using a tape measure, with marker cones placed at each end to indicate running distance.

• Participant shall be equipped with the Cortex Metamax 3B gas analysis equipment and a heart rate monitor.

• Participant is required to continuously run between the two cones, 15-meter distance, in time with the beeps from the audio recording. Upon failing to reach the distance within the beep, the participant is required to speed up to reach the cone in time with the beep. Three consecutive fails shall result in the test being stopped.

• The test begins at a steady speed and gradually increases (time between beeps decreases) as the participant progresses through each level of the test.

• Researchers shall calibrate the RPE scale with participants prior to the start of the test and also verbally explain the procedure and provide them with a chance to listen to the audio recording.

• The number of completed shuttles shall be recorded; RPE and HR shall be recorded at the end of each stage of the test and at each level specified in Table 1.

• Participants shall run until either they reach volitional exhaustion/wish to stop the test. The test will be stopped by researchers if RPE is recorded at 18 or above.

Appendix F.

Anthropometric measurements

Stature (Height) Measurement

- The participant's height will be measured to the nearest 0.1cm using a wall stadiometer (Seca, Germany and Tanita, USA).
- The participant is instructed to stand underneath the stadiometer facing away from the wall, with heels, scapulae and buttocks in contact with the wall.
- Participant will be asked to stand as tall as possible with heels together and feet evenly balanced at approximately 60.
- The participant will be asked to inhale deeply and maintain the position; one edge of the sliding scale will be lowered onto the participants head and the participant then asked to step away.
- The height to the nearest tenth of a centimetre (0.1cm) will then be recorded.

Mass Measurement

Participants body mass will be measured to the nearest 0.1kg using an electronic scale (Seca, Germany). The results from the measurement of stature and mass will be used to determine the participants body mass index (BMI) using the formula BW(kg)/Ht2(M).

Appendix G.

University of Chester

Blood Pressure measurement

Equipment: Omron (BP710) automated blood pressure device.

1. Participant is asked to remove tight-fitting clothing from the upper arm.

2. The cuff is applied to the left upper arm, so the arrow is cantered on the inside of the arm and aligned with the middle finger. The air tube runs down the inside of the participants arm. The bottom of the cuff should be approximately ½” above the elbow.

3. Participant is asked to sit on a chair with their feet flat on the floor.

4. The arm is placed on a table, so the cuff is level with their heart.

5. Participant is asked to keep still and not to talk during measurement.

6. Press the START/STOP button.

7. The cuff starts to inflate automatically. As the cuff inflates, the monitor automatically determines the ideal inflation level.

8. Participant is asked to remain still and not move their arm until the entire measurement process is completed.

9. Inflation stops automatically, and the measurement is started. As the cuff deflates, decreasing numbers appear on the display and the Heartbeat Symbol will flash.

10. When the measurement is complete, the arm cuff completely deflates. The blood pressure and pulse rate are displayed.

Appendix H.

<u>Gas analysis and HR telemetry equipment</u>

Cortex Metamax 3B will be used to measure oxygen uptake (**$\dot{V}O_2$**) during CTWT, CTT and 15m MSFT. The device weighs <1kg and can be carried on the front or back using the carrying system.

http://www.cortex-medical.de/METAMAX-3B-en.htm

Polar HR monitor will be used to transmit HR data directly to the Cortex Metamax 3B during exercise testing. This strap is worn around the chest of the participant during exercise.

http://www.polar.com/uk-en/products

Appendix I.

Additional data

Supplementary Table 1. Trial two descriptive statistics for $\dot{V}O_2$, RER, %HRMAX and %HRR.

Level	Estimated Aerobic Capacity ($ml.kg^{-1}.min^{-1}$)	$\dot{V}O_2$ ($ml.kg^{-1}.min^{-1}$)	RER	%HRmax	% HRR
5:4	35	40.7±4.7*	0.98±0.07	88±7	81±10
7:6	41	46.7±6.3*	1.06±0.07	93±5	89±8
9:4	46	51.7±5.2*	1.12±0.09	95±4	93±5
10:5	51	54.5±6.3*	1.15±0.08	97±4	95±6

*Significant difference from predicted $\dot{V}0_2$.

Supplementary Table 2. Descriptive statistics for the paired samples t-test for Trial 1 and Trial 2 for Oxygen uptake ($\dot{V}O_2$).

Levels	Number of participants	Trial 1: Measured $\dot{V}O_2$ ($ml \cdot kg^{-1} \cdot min^{-1}$) (95% CI)	Trial 2: Measured $\dot{V}O_2$ ($ml \cdot kg^{-1} \cdot min^{-1}$) (95% CI)	Mean Difference between trials (Δ)
5:4	43	40.4±5.3 (38.7-40.0)	40.8±4.9 39.3-42.3	-0.41 (-2.01-1.20)
7:6	36	46.3±5.9 (44.3-48.3)	46.9±6.4 (44.7-49.1)	-0.57 (-1.95-0.81)
9:4	15	52.0±7.9 (47.7-56.4)	52.3±5.0 (49.5-55.1)	-0.29 (-2.46-1.87)
10:5	7	55.2±7.5 (48.3-62.1)	53.7±6.9 (47.4-60.1)	1.51 (-4.91-7.93)

Supplementary Table 3. Descriptive statistics for the paired samples t-test for Trial 1 and Trial 2 for Respiratory Exchange Ratio (RER).

Levels	**Trial 1: RER (95% CI)**	**Trial 2: RER (95% CI)**	**Mean Difference between trials (Δ) (95% CI)**
5:4	1.00±0.09 (38.7-40.0)	0.98±0.07 39.3-42.3	0.02 (-0.02-0.05)
7:6	1.08±0.08 (44.3-48.3)	1.05±0.08 (44.7-49.1)	0.03 (0.00-0.05)
9:4	1.15±0.10 (47.7-56.4)	1.12±0.10 (49.5-55.1)	0.03 (-0.04-0.10)
10:5	1.16±0.13 (48.3-62.1)	1.12±0.11 (47.4-60.1)	0.36 (-0.10-0.18)